

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:18:10 ; Search time 30.04 seconds  
(without alignments)  
410.426 Million cell updates/sec

Title: US-09-662-784-6

Perfect score: 111

Sequence: 1 DTMGALLVLLVLTQALGV.....CMGEAVONTVEDKINTLGR 111

Scoring table:

OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	100.0	111	13	TRFP Chain #2 with
2	111	100.0	111	21	T cell reactive fe
3	111	100.0	111	21	Feline human TRFP
4	111	100.0	111	21	Cat TRFP chain 2 1
5	111	100.0	111	21	Human TRFP chain 2
6	109	98.2	109	14	Human TRFP chain 2 (with
7	109	98.2	109	14	Human T cell react
8	109	98.2	109	20	Felis sp. allergen
9	99	89.2	111	12	TRFP chain 2 - lon
10	92	82.9	92	21	T cell reactive fe
11	92	82.9	92	21	Feline human TRFP

12	92	82.9	92	21	AAV51475	Human TRFP chain 2
13	83	74.8	97	12	AA12123	TRFP chain 2 - tru
14	83	74.8	109	21	AA12123	T cell reactive fe
15	83	74.8	109	21	AA12123	Feline human TRFP
16	83	74.8	109	21	AA12123	Cat TRFP chain 2 s
17	83	74.8	109	21	AA12123	Human TRFP chain 2
18	83	74.8	110	13	AA12123	TRFP chain #2 with
19	81	73.0	97	21	AA12123	T cell reactive fe
20	81	73.0	97	21	AA12123	Feline human TRFP
21	81	73.0	97	21	AA12123	Cat TRFP chain 2 t
22	81	73.0	97	21	AA12123	Human TRFP chain 2
23	79	71.2	109	12	AA12123	TRFP I chain 2 - s
24	66	59.5	92	18	AA12123	Chain 2 of major c
25	66	59.5	92	20	AA12123	Human MHC Class II
26	64	57.7	90	21	AA12123	T cell reactive fe
27	64	57.7	90	21	AA12123	Feline human TRFP
28	64	57.7	90	21	AA12123	Human TRFP chain 2
29	51	45.9	101	13	AA12123	TRFP chain #2 C2ST
30	37	33.3	51	21	AA12123	Feline human TRFP
31	37	33.3	51	21	AA12123	Human TRFP chain 2
32	37	33.3	51	21	AA12123	TRFP chain #2 C2ST
33	37	33.3	82	13	AA12123	T cell reactive fe
34	32	28.8	82	21	AA12123	Feline human TRFP
35	32	28.8	82	21	AA12123	Cat TRFP chain 2 g
36	32	28.8	82	21	AA12123	Human TRFP chain 2
37	32	28.8	82	21	AA12123	Peptide Z. Felis
38	26	23.4	26	14	AA12123	Human T cell react
39	26	23.4	26	14	AA12123	Peptide Z derived
40	26	23.4	26	21	AA12123	Feline human TRFP
41	26	23.4	26	21	AA12123	Cat TRFP derived p
42	26	23.4	26	21	AA12123	Human TRFP derived
43	26	23.4	26	14	AA12123	Recombinant YZX. Fel
44	26	23.4	26	14	AA12123	Peptide YZX. Fel
45	26	23.4	26	21	AA12123	Peptide YZX. Fel

#### ALIGNMENTS

RESULT 1	
AA12123	standard; protein; 111 AA.
XX	
AC	AA12123;
XX	
DT	20-MAY-1998 (first entry)
XX	
DE	TRFP Chain #2 with C2 leader sequence.
XX	
KW	T cell reactive feline protein; cat allergy; allergic; Ige;
XX	
OS	Felis domesticus.
XX	
FT	key
FT	peptide
FT	1..19
FT	/label= C2 leader
FT	19..111
FT	/label= TRFP chain #2
XX	
PN	WO9215613-A.
XX	
PD	17-SEP-1992.
XX	
PF	20-FEB-1992; 92WO-0501344.
XX	
PR	28-FEB-1991; 91US-0662193.
XX	
PA	(IMMU-) IMMUNOLOGIC PHARM CORP.
XX	
PI	Bond J, Kuo M;
XX	
DR	WPI, 1992-331670/40.



CC different covalently linked peptide chains with a molecular weight of 20  
 CC KD, 40 KD or 130 KD under non-reducing conditions and 5 KD or 10-18 KD  
 CC under reducing conditions. The products of the invention have  
 CC anti-allergic activity and act as human T cell stimulators. TRPP is useful  
 CC for reducing or preventing the adverse effects of cat allergens on cat  
 CC allergic individuals and in ex vivo diagnostic tests to determine which  
 CC peptides cause sensitivity so as to selectively use them to desensitize  
 CC a cat sensitive individual. Purified TRPP is also useful for studying  
 CC the mechanism of immunotherapy of cat allergy and to design modified  
 CC derivatives, analogs or functional equivalents that are more useful in  
 CC immunotherapy against cat allergy. DNA sequences encoding TRPP are  
 CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
 CC and/or therapeutics. Fully defined and characterized TRPP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d I)  
 CC chain 2, long form which is described in the method of the invention.

XX Sequence 111 AA:

SO

Query Match 100.0%; Score 111; DB 21; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-100;  
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTRMGALLVALLVYQALGYKMAETCPIFYDVFPVANGNELLDLSLTKVNATEPERTA 60  
 Db 1 dtrmgallvallaivtqalgymaetcpifdvffavangnellldsltkvnapertta 60

OY 61 MKRIQDCYVENGISRVLGIVMTTSSSKDCMGCAVQNTVEDIKNTLGR 111  
 Db 61 mkrigdcyvenslrsrldglvmttssskdcmgcaqvntvedikntlgr 111

RESULT 4  
 AAY90103 standard; Protein: 111 AA.

XX AAY90103;

AC

DT 13-JUL-2000 (first entry)

XX

DE Cat TRFP chain 2 long form protein sequence.

XX

KM Cat: TRFP: human T-cell reactive feline protein; cat protein allergen;  
 KM house dust; Fel d I; cat allergy; Fells domesticus sensitivity; therapy;  
 KM diagnosis; goat; sheep; horse; rabbit; dog.

XX

OS Fells domesticus.

XX

XX Key Location/Qualifiers  
 FH Peptide 1..19  
 FT /note- "signal peptide"  
 FT 20..111  
 FT Protein /note- "mature TRFP chain 2 long form"

XX

PN US6025162-A.

XX

PD 15-FEB-2000.

XX

PF 28-APR-1995; 95US-0430944.

XX

XX 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.

XX

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Morgenstern JP, Griffith ID, Rogers BL;

XX  
 DR WPI; 2000-181812/16.  
 DR N-PSDB; AAA07437.

XX  
 PT New human T cell reactive feline protein, useful for desensitizing cat  
 XX allergic individuals to cat allergens -

PS Claim 1; Fig 3; 108pp; English.

XX  
 CC This sequence is a peptide chain of the human T cell reactive feline  
 CC protein (TRPP) of the invention. The protein is a cat protein allergen,  
 CC and was isolated from a vacuum bag extract obtained by affinity  
 CC purification of house dust collected from several homes with cats. TRPP  
 CC is composed of two covalently linked peptide chains, and is also referred  
 CC to as Fel d I. TRPP and its peptides are useful for reducing or  
 CC preventing the adverse effects that exposure to cat allergens normally  
 CC has on cat allergic individuals (i.e. to desensitize individuals to cat  
 CC allergens or block the effect of the allergens). TRPP is also used in  
 CC methods of diagnosing sensitivity to Fells domesticus in an individual.  
 CC DNA sequences encoding TRPP can be used as probes to locate equivalent  
 CC sequences present in other species, e.g. goat, sheep, horse, rabbit and  
 CC dog, that may be useful in a diagnostic and/or therapeutic applications.

XX  
 SO Sequence 111 AA:

Query Match 100.0%; Score 111; DB 21; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-100;  
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTRMGALLVALLVYQALGYKMAETCPIFYDVFPVANGNELLDLSLTKVNATEPERTA 60  
 Db 1 dtrmgallvallaivtqalgymaetcpifdvffavangnellldsltkvnapertta 60

OY 61 MKRIQDCYVENGISRVLGIVMTTSSSKDCMGCAVQNTVEDIKNTLGR 111  
 Db 61 mkrigdcyvenslrsrldglvmttssskdcmgcaqvntvedikntlgr 111

RESULT 5  
 AAY51470 standard; Protein: 111 AA.

XX AAY51470;

AC

DT 22-MAY-2000 (first entry)

XX

DE Human TRPP chain 2 (long form) protein fragment.

XX

KM T-cell reactive feline protein; TRPP; T cell epitope; T cell receptor;  
 KM down regulation; immune response; allergen; immunoglobulin E;  
 KM sensitivity; cat protein allergen; human; chain 2.

XX

OS Homo sapiens.

XX

PN US6019772-A.

XX

PD 01-FEB-2000.

XX

PF 02-SEP-1994; 94US-0300928.

XX

XX 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.

XX

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;  
 XX WPI; 2000-146862/13.

DR N-PSDB; AA288617.  
XX  
XX  
PT Peptides of human T cell reactive feline protein for treating  
PT sensitivity to cat protein allergens comprise at least one T cell  
PT epitope recognized by a T cell receptor specific for the human T cell  
PT reactive feline protein -  
XX  
XX  
PS Claim 1: Column 75-78; 105pp; English.  
CC This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94,  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC chain 2 (long form).  
XX  
XX  
SQ Sequence 111 AA:  
  
Query Match 100.0%; Score 111; DB 21; Length 111;  
Best Local Similarity 100.0%; Pred. No. 3.3e-100;  
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DTRGALLVALLVTOALGVKAEETCFYDVFPAVANGNELLDLSITRYNATEPERAMK 60  
DB 1 dtrmgallviallvlgalgkmaetcpfydvffavangnelldlsitrynatpeperla 60  
QY 61 MKKIQDCYVENGISRYLDGLVMTTSSSKDCGSAVQNTVEDLKLNTLGR 111  
DB 61 mkkidcyvenglisryldglvmttssskdcgseavqntvedlklntlgr 111  
  
RESULT 6  
ID AAR36541 standard; Protein; 109 AA.  
AC AAR36541;  
XX  
XX  
DT 12-AUG-1993 (first entry)  
XX  
XX  
DE TRFP chain 2 (with leader).  
XX  
XX  
KW Human T cell reactive feline protein; TRFP; leader A; leader B;  
KW epitope.  
XX  
XX  
OS Felis.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..17  
FT /label= leader-peptide  
XX  
XX  
PN WO9308280-A.  
XX  
XX  
PD 29-APR-1993.  
XX  
XX  
PF 16-OCT-1992; 92WO-US08694.  
XX  
XX  
PR 16-OCT-1991; 91US-0777859.  
PR 13-DEC-1991; 91US-0807529.  
XX  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
XX  
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;  
PI Rogers BL.  
XX  
XX  
DR WPI: 1993-152473/18.  
DR N-PSDB; AAQ41558.  
XX

PT Recombitope peptide having T-cell stimulating activity - for the  
PT diagnosis and treatment of sensitivity to protein allergens,  
PT auto-antigens and protein antigens  
XX  
XX  
PS Disclosure; Fig 2; 73pp; English.  
CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.  
CC coli and purified. T cell epitope studies using overlapping peptide  
CC regions derived from the TRFP amino acids sequence were used to  
CC identify multiple T cell epitopes in each chain of TRFP.  
XX  
XX  
SQ Sequence 109 AA:  
  
Query Match 98.2%; Score 109; DB 14; Length 109;  
Best Local Similarity 100.0%; Pred. No. 3e-98;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 MRGALLVALLVTOALGVKAEETCFYDVFPAVANGNELLDLSITRYNATEPERAMK 62  
DB 1 mrgallviallvlgalgkmaetcpfydvffavangnelldlsitrynatpeperamk 60  
QY 63 KIDQCYVENGISRYLDGLVMTTSSSKDCGSAVQNTVEDLKLNTLGR 111  
DB 61 kidqcyvenglisryldglvmttssskdcgseavqntvedlklntlgr 109  
  
RESULT 7  
ID AAR41985 standard; Protein; 109 AA.  
AC AAR41985;  
XX  
XX  
DT 21-APR-1994 (first entry)  
XX  
XX  
DE Human T cell reactive feline protein chain 2.  
XX  
XX  
KW Human; T cell; reactive; feline; protein; immune response; antigen;  
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Caris;  
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemisia;  
KW Plantago; Parietaria; Blatella; Apis; Periplaneta; autoantigen.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..17  
FT /note= "Signal peptide"  
FT Protein 18..109  
FT /note= "Mature protein"  
XX  
XX  
PN WO9319178-A.  
XX  
XX  
PD 30-SEP-1993.  
XX  
XX  
PF 25-MAR-1993; 93WO-US02462.  
XX  
XX  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
XX  
PI Briher TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;  
PI Morville M.  
XX  
XX  
DR WPI: 1993-320744/40.  
DR N-PSDB; AAQ49535.  
XX  
XX  
PT New peptide(s) for inducing tolerance - comprise one or more  
PT epitope(s) of an allergen administered subcutaneously, for  
PT treating sensitivity to cats, bees, etc.  
XX  
XX  
PS Disclosure; Fig 2; 107pp; English.



XX This sequence represents chain 2 of human T cell reactive feline  
 CC protein (TRFP). Peptides derived from TRFP may be used in a  
 CC therapeutic composition which is useful in treating diseases which  
 CC involve an immune response to a protein antigen. This composition  
 CC may be used to induce tolerance in a mammal to dermatophagoides,  
 CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,  
 CC Quercus, Olea, Artemisia, Plantago, Parietaria, Canis, Blatella,  
 CC Apis, Periplaneta and to autoantigens in humans.

XX Sequence 109 AA:

Query Match 98.2%; Score 109; DB 14; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 3e-98;  
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 MEGALLVALVLTQALGVKAETCPFYDYFFAVANGNELLDLSLTKVNATEPERTAMK 62  
 Db 1 mrgallvalvltqalgvknaetcpfydyffavangnellldsltkvnatepertamk 60  
 OY 63 KIDCVYENGLISRVLDGLVMTTSSSKDCMGEAVONTVEDLKINTLGR 111  
 Db 61 kidcvyenglisrvldglvmttssskdcmgavontvedlkinltlgr 109

#### RESULT 8

AAV25593  
 ID AAV25593 standard; protein; 109 AA.

XX AAV25593;

DT 30-SEP-1999 (first entry)

DE Felis sp. allergen 1082946 Fel di chain 2 precursor protein fragment.

XX Major histocompatibility complex; class II; desensitizing; human;  
 KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Felis sp.

XX WO9934826-A1.

XX 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB000080.

XX 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

DR WPI; 1999-458255/38.

PT Desensitizing patients to polypeptide allergens

XX Example 6; Page 53; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a  
 CC polypeptide allergen and comprises administering to the patient a peptide  
 CC derived from the allergen where restriction to a MHC Class II molecule  
 CC possessed by the patient can be demonstrated for the peptide and the  
 CC peptide is able to induce a late phase response in an individual who  
 CC possesses the MHC Class II molecule. The methods can be used for  
 CC desensitizing patients to allergens present in e.g. grass, tree and weed  
 CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit

CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 CC tentorial mottor beetle, mammals such as cat, dog, horse, cow, pig,  
 CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 CC produce immunological vaccines which may be used to prevent and/or treat  
 CC conditions involving hypersensitivity to allergens. This sequence  
 CC represents the cat allergen 1082946 Fel di chain 2 precursor.

SO Sequence 109 AA:

Query Match 98.2%; Score 109; DB 20; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 3e-98;  
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 MEGALLVALVLTQALGVKAETCPFYDYFFAVANGNELLDLSLTKVNATEPERTAMK 62  
 Db 1 mrgallvalvltqalgvknaetcpfydyffavangnellldsltkvnatepertamk 60  
 OY 63 KIDCVYENGLISRVLDGLVMTTSSSKDCMGEAVONTVEDLKINTLGR 111  
 Db 61 kidcvyenglisrvldglvmttssskdcmgavontvedlkinltlgr 109

#### RESULT 9

AA12121  
 ID AA12121 standard; protein; 111 AA.

XX AA12121;

DT 26-JUL-1991 (first entry)

DE TRFP chain 2 - long form.

XX Human T cell reactive feline protein; cat allergens.

XX Felis catus.

FH Key location/Qualifiers

FT Peptide 3..19 /label= Leader sequence

FT Protein 20..111 /label= TRFP I long form

XX WO9106571-A.

XX 16-MAY-1991

PF 02-NOV-1990; 90WO-US06548.

XX 03-NOV-1989; 89US-0431565.

XX (IMMU-) IMMUNOLOGIC PHARM COR.

PI Gelfer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;

DR WPI; 1991-164136/22.

DR N-PSDB; AAQ11838.

PT New pure covalently linked human T cell reactive feline protein

XX and to diagnose sensitivity to allergens.

XX Claim 2; Fig 3; 70pp; English.

XX poly-A mRNA from cat parotid and mandibular glands was used to  
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These  
 CC clones were then used to screen a cat genomic library. Chain 1  
 CC exists in two forms having different leader sequences (A and B).  
 CC The sequence can be used to express the protein and peptide derivs.  
 CC which stimulate T-cells in persons allergic to cats. The peptides  
 CC can be used to reduce/eliminate the allergic response partic. by

CC modifcn. of lymphokine prodn. by the T-cells. They can also be  
CC used to identify epitopes responsible for sensitivity. The DNA can  
CC be used to detect comparable sequence in other species, and also  
CC for prodn. of modified forms of TRPP esp. showing reduced binding  
CC to IgE and thus reduced tendency to cause adverse reactions.  
CC See also AARI2119-R12123.

SO Sequence 111 AA:

Query Match 89.28; Score 99; DB 12; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.0e-88;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFMGALVIALVIAVQAGVMAETCPITFDVFPVANGNELLDLSITKVNATEPERTA 60  
DB 1 dcmgallviallvicqalgvmaetcpitfdvffavangnellldlsitkvnateperta 60

OY 61 MKKIQCCEYBNGISRVLDGLVMTTSSSKDCMGEAVON 99  
DB 61 mkkiqcceyenglsrvldglvmttssskdcmgavon 99

RESULT 10  
AAB28937

ID AAB28937 standard; Protein; 92 AA.

AC AAB28937;

DT 29-JAN-2001 (first entry)

DE T cell reactive feline protein chain 2 PRO long.

KW Cat; allergy; human T cell reactive feline protein; hTRFP;  
KW immunotherapy.

OS Felis sp.

PV US6120769-A.

PD 19-SEP-2000.

PE 28-APR-1995; 95US-0431184.

PR 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 23-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Geffer ML, Garman RD, Greenstein JL, Bond JF;

PT WPI; 2000-601477/57.

PS Detecting, preventing and treating sensitivity to cat protein allergen  
PS comprises combining a biological sample with a human T cell reactive  
PS feline protein and determining the extent of binding that occurs -

PS Disclosure; Figure 7; 106pp: English.

CC The present invention relates to the detection of sensitivity to a cat  
CC protein allergen by combining a blood sample from a subject with a  
CC peptide of human T cell reactive feline protein (hTRFP). This method  
CC and the hTRFP peptides are useful for diagnosing, preventing and  
CC treating cat allergies by reducing or abolishing an individual's  
CC allergic response to a cat allergen. DNA encoding the TRFP may be  
CC used as probes to locate equivalent sequences present in other species.  
CC These may further be used to study the mechanism of immunotherapy of  
CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present  
CC sequence was used in the invention.

SO Sequence 92 AA:

Query Match 82.98; Score 92; DB 21; Length 92;  
Best Local Similarity 100.0%; Pred. No. 8.8e-82;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 VKMAETCPITFDVFPVANGNELLDLSITKVNATEPERTAMKIIDCEVNGLSRVLD 79  
DB 1 vkmaetcpitfdvffavangnellldlsitkvnatepertamkkidcyvenglsrvld 60

OY 80 GLVMTTSSSKDCMGEAVONTVEDLKLNTLGR 111  
DB 61 glvmttssskdcmgavontvedlklntlgr 92

RESULT 11  
AAV87678

ID AAV87678 standard; Protein; 92 AA.

AC AAV87678;

DT 22-AUG-2000 (first entry)

DE Feline human TRFP chain 2 Long form protein #2.

KW T-cell reactive feline protein; TRFP; Fel d 1; cat allergen;  
KW antiallergic; T cell stimulator; diagnostic; immunotherapy.

OS Felis sp.

PV US6048962-A.

PD 11-APR-2000.

PE 27-APR-1995; 95US-0430014.

PR 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;  
PI Greenstein JL, Griffith ID, Garman RD;

PT WPI; 2000-316905/27.

PS New human T cell reactive feline protein useful for reducing or  
PS abolishing individual's allergic response to cat allergen comprising  
PS two different covalently linked peptide chains -

PS Example 2; Column 81-84; 106pp: English.

CC This invention describes a novel naturally occurring cat protein allergen  
CC (I), human T cell reactive feline protein (TRFP), comprising two  
CC different covalently linked peptide chains with a molecular weight of 20  
CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
CC under reducing conditions. The products of the invention have  
CC antiallergic activity and act as human T cell stimulators. TRFP is useful  
CC for reducing or preventing the adverse effects of cat allergens on cat  
CC allergic individuals and in ex vivo diagnostic tests to determine which  
CC peptides cause sensitivity so as to selectively use them to desensitize  
CC a cat sensitive individual. Purified TRFP is also useful for studying  
CC the mechanism of immunotherapy of cat allergy and to design modified  
CC derivatives, analogs or functional equivalents that are more useful in

CC Immunotherapy against cat allergy. DNA sequences encoding TRFP are  
 CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
 CC and/or therapeutics. Fully defined and characterized TRFP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d 1)  
 CC chain 2, long form which is described in the method of the invention.

XX  
 SQ Sequence 92 AA;

Query Match 82.9%; Score 92; DB 21; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-82;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 VKAETCPITYDVEFAVANGNELLDLSLRKVNATEPERTAMRKIDQCYENGISVLD 79  
 Db 1 vkmaetcpitydvffavangnellldlsrkvatepertamrkidqcyengisvld 60  
 OY 80 GLVWTTTSSSKDCMGEAVQNTVEDLKNTLGR 111  
 Db 61 glvmttssskdcmgeavqntvedlknltlgr 92

RESULT 12  
 AAY51475  
 ID AAY51475 standard; Protein; 92 AA.  
 AC AAY51475;  
 DT 22-MAY-2000 (first entry)

DE Human TRFP chain 2 (long form) protein fragment #2.  
 XX  
 KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
 KM down regulation; immune response; allergen; immunoglobulin E;  
 XX sensitivity; cat protein allergen; human; chain 2.

OS Homo sapiens.  
 XX  
 PN US6019972-A.  
 PD 01-FEB-2000.  
 PF 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;  
 XX WPI; 2000-146862/13.

DR Peptides of human T cell reactive feline protein for treating  
 XX  
 PT sensitivity to cat protein allergens comprise at least one T cell  
 PT epitope recognized by a T cell receptor specific for the human T cell  
 PT reactive feline protein -

XX  
 PS Example 1; Column 83-84; 105pp; English.

XX This invention describes a novel peptide (I) of human T cell reactive  
 CC feline protein (hTRFP) having at least one T cell epitope recognized  
 CC by a T cell receptor specific for the human T cell reactive feline  
 CC protein, the peptide consisting of at least 7-30 amino acids, and having  
 CC an amino acid sequence derived from an amino acid sequence comprising 94,  
 CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
 CC down regulate the immune response to the allergen. The peptides have

CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
 CC peptide (I) is useful in compositions for treating sensitivity to a cat  
 CC protein allergen in a subject. This sequence represents the human TRFP  
 CC chain 2 (long form).

XX  
 SQ Sequence 92 AA;

Query Match 82.9%; Score 92; DB 21; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-82;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 VKAETCPITYDVEFAVANGNELLDLSLRKVNATEPERTAMRKIDQCYENGISVLD 79  
 Db 1 vkmaetcpitydvffavangnellldlsrkvatepertamrkidqcyengisvld 60  
 OY 80 GLVWTTTSSSKDCMGEAVQNTVEDLKNTLGR 111  
 Db 61 glvmttssskdcmgeavqntvedlknltlgr 92

RESULT 13  
 AAR12123  
 ID AAR12123 standard; Protein; 97 AA.  
 AC AAR12123;  
 DT 26-JUL-1991 (first entry)  
 DE TRFP chain 2 - truncated short form.

KW Human T cell reactive feline protein; cat allergens.

OS Felis catus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 3..19  
 FT /label= Leader sequence  
 FT Protein 20..97  
 FT /label= TRFP I truncated short form

PN WO9106571-A.  
 PD 16-MAY-1991.  
 PF 02-NOV-1990; 90WO-US06548.  
 PR 03-NOV-1989; 89US-0431565.  
 PA (IMMU-) IMMUNOLOGIC PHARM COR.

XX Gelfer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;  
 PI Brauer AW;  
 DR WPI; 1991-164136/22.  
 DR N-PSDB; AAQ11840.

PT New pure covalently linked human T cell reactive feline protein -  
 PT and modified peptide(s), used to reduce effects of cat allergens  
 PT and to diagnose sensitivity to allergens.

XX  
 PS Claim 2; Fig 5; 70pp; English.

XX Poly-A mRNA from cat parotid and mandibular glands was used to  
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These  
 CC clones were then used to screen a cat genomic library. Chain 1  
 CC exists in two forms having different leader sequences (A and B).  
 CC The sequence can be used to express the protein and peptide derivs.  
 CC which stimulate T-cells in persons allergic to cats. The peptides  
 CC can be used to reduce/eliminate the allergic response partic. by  
 CC modifcn. of lymphokine prodn. by the T-cells. They can also be  
 CC used to identify epitopes responsible for sensitivity. The DNA can  
 CC be used to detect comparable sequence in other species, and also

CC for prodn. of modified forms of TRFP esp. showing reduced binding  
 CC to IGE and thus reduced tendency to cause adverse reactions.  
 CC See also AAR12119-R12122.  
 XX

SQ Sequence 97 AA:

Query Match 74.8%; Score 83; DB 12; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 5,2e-73;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFMGALLVALLVTOALGVKMAETCPIDYDFVFAVANGNELLDLSLRKNATEPERTA.60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 dnmrgallvllalvtqalgykmaetcpilfydvffavangnellldlsitkvnateperta.60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 61 MKRIQCYVENGILSRVLDGLV 83  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 mkkldcyvengilslrvldglvm 83  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

#### RESULT 14

AAB28934  
 ID AAB28934 standard; Protein; 109 AA.

XX AC AAB28934:

XX DT 29-JAN-2001 (first entry)

XX DE T cell reactive feline protein chain 2 short form.

XX KN Cat: allergy; human T cell reactive feline protein; hTRFP;  
 immunotherapy.

XX OS Felis sp.

XX PN US6120769-A.

XX PD 19-SEP-2000.

XX PF 28-APR-1995; 95US-0431184.

XX PR 02-SEP-1994; 94US-0300928.

XX PR 03-NOV-1989; 89US-0431565.

XX PR 28-FEB-1991; 91US-0662276.

XX PR 13-DEC-1991; 91US-0807529.

XX PR 25-MAR-1992; 92US-0857311.

XX PR 15-MAY-1992; 92US-0884718.

XX PR 15-JAN-1993; 93US-0006116.

XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.

XX PI Geffer ML, Garman RD, Greenstein JL, Bond JP;

XX DR WPI: 2000-601477/57.

XX DR N-PSDB: AAC50104.

XX PT Detecting, preventing and treating sensitivity to cat protein allergen  
 PT comprises combining a biological sample with a human T cell reactive  
 PT feline protein and determining the extent of binding that occurs -

XX PS Claim : Figure 4; 106pp; English.

XX The present invention relates to the detection of sensitivity to a cat  
 CC protein allergen by combining a blood sample from a subject with a  
 CC peptide of human T cell reactive feline protein (hTRFP). This method  
 CC and the hTRFP peptides are useful for diagnosing, preventing and  
 CC treating cat allergies by reducing or abolishing an individual's  
 CC allergic response to a cat allergen. DNA encoding the TRFP may be  
 CC used as probes to locate equivalent sequences present in other species.  
 CC These may further be used to study the mechanism of immunotherapy of  
 CC cat allergy, and to design modified derivatives, analogues or  
 CC functional equivalents useful in immunotherapy. The present  
 CC sequence was used in the invention.

XX SQ Sequence 109 AA:

Query Match 74.8%; Score 83; DB 21; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 5,9e-73;

Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DFMGALLVALLVTOALGVKMAETCPIDYDFVFAVANGNELLDLSLRKNATEPERTA.60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 dnmrgallvllalvtqalgykmaetcpilfydvffavangnellldlsitkvnateperta.60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 61 MKRIQCYVENGILSRVLDGLV 83  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 mkkldcyvengilslrvldglvm 83  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

#### RESULT 15

AAY87674  
 ID AAY87674 standard; Protein; 109 AA.

XX AC AAY87674:

XX DT 22-AUG-2000 (first entry)

XX DE Feline human TRFP chain 2 short form protein.

XX KN T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
 KW anti-allergic; T cell stimulator; diagnostic; immunotherapy.

XX OS Felis sp.

XX PN US6048962-A.

XX PD 11-APR-2000.

XX PF 27-APR-1995; 95US-0430014.

XX PR 02-SEP-1994; 94US-0300928.

XX PR 03-NOV-1989; 89US-0431565.

XX PR 28-FEB-1991; 91US-0662276.

XX PR 13-DEC-1991; 91US-0807529.

XX PR 25-MAR-1992; 92US-0857311.

XX PR 15-MAY-1992; 92US-0884718.

XX PR 15-JAN-1993; 93US-0006116.

XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.

XX PI Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AM;

XX DR Greenstein JL, Griffith TJ, Garman RD;

XX DR WPI: 2000-316905/27.

XX DR N-PSDB: AAA12245.

XX PT New human T cell reactive feline protein useful for reducing or  
 PT abolishing individual's allergic response to cat allergen comprising  
 PT two different covalently linked peptide chains -

XX PS Claim 2; Column 77-78; 106pp; English.

XX This invention describes a novel naturally occurring cat protein allergen  
 CC (1), human T cell reactive feline protein (TRFP), comprising two  
 CC different covalently linked peptide chains with a molecular weight of 20  
 CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
 CC under reducing conditions. The products of the invention have  
 CC anti-allergic activity and act as human T cell stimulators. TRFP is useful  
 CC for reducing or preventing the adverse effects of cat allergens on cat  
 CC allergic individuals and in ex vivo diagnostic tests to determine which  
 CC peptides cause sensitivity so as to selectively use them to desensitize  
 CC a cat sensitive individual. Purified TRFP is also useful for studying  
 CC the mechanism of immunotherapy of cat allergy and to design modified  
 CC derivatives, analogs or functional equivalents that are more useful in  
 CC immunotherapy against cat allergy. DNA sequences encoding TRFP are

CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
 CC and/or therapeutics. Fully defined and characterized TRFP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d 1)  
 CC chain 2, short form which is described in the method of the invention.

SO Sequence 109 AA:

Query Match 74.8%; Score 83; DB 21; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-73;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTMRGALIVLALVTQALGVKMAETCPFYDVFPAVANGNELLDLSLTQVNAPEPRTA 60  
 Db 1 dcmgallivallvtqalgvkmaetcpfydvifavangnellldsltkvnateperta 60

OY 61 MKRIQDCYVENGSLISRYLDGLVM 83  
 Db 61 mkriqdcyvenngslisryldglvm 83

RESULT 16  
 AAY90104  
 ID AAY90104 standard; Protein; 109 AA.

XX AAY90104;

DT 13-JUL-2000 (first entry)

DE Cat TRFP chain 2 short form protein sequence.

KM Cat; TRFP: human T-cell reactive feline protein; cat protein allergen;  
 KM house dust; Fel d 1; cat allergy; Felis domesticus sensitivity; therapy;  
 KM diagnosis; goat; sheep; horse; rabbit; dog.

OS Felis domesticus.

FT Key Location/Qualifiers

FT Peptide 1..19 /note- "signal peptide"

FT Protein 20..111 /note- "mature TRFP chain 2 short form"

PN US6025162-A.

PD 15-FEB-2000.

PF 28-APR-1995; 95US-0430944.

PR 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (TMMD-) IMMUNOLOGIC PHARM CORP.

PI Morgenstern JP, Griffith IJ, Rogers BL;

DR WPI: 2000-181812/16.

DR N-PSDB; AAA07436.

XX New human T cell reactive feline protein, useful for desensitizing cat  
 XX allergic individuals to cat allergens -  
 XX Claim 1; Fig 4; 108bp; English.

CC This sequence is a peptide chain of the human T cell reactive feline  
 CC protein (TRFP) of the invention. The protein is a cat protein allergen,

CC and was isolated from a vacuum bag extract obtained by affinity  
 CC purification of house dust collected from several homes with cats. TRFP  
 CC is composed of two covalently linked peptide chains, and is also referred  
 CC to as Fel d 1. TRFP and its peptides are useful for reducing or  
 CC preventing the adverse effects that exposure to cat allergens normally  
 CC has on cat allergic individuals (i.e. to desensitize individuals to cat  
 CC allergens or block the effect of the allergens). TRFP is also used in  
 CC methods of diagnosing sensitivity to Felis domesticus in an individual.

CC DNA sequences encoding TRFP can be used as probes to locate equivalent  
 CC sequences present in other species, e.g. goat, sheep, horse, rabbit and  
 CC dog, that may be useful in a diagnostic and/or therapeutic applications.

SO Sequence 109 AA:

Query Match 74.8%; Score 83; DB 21; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-73;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTMRGALIVLALVTQALGVKMAETCPFYDVFPAVANGNELLDLSLTQVNAPEPRTA 60  
 Db 1 dcmgallivallvtqalgvkmaetcpfydvifavangnellldsltkvnateperta 60

OY 61 MKRIQDCYVENGSLISRYLDGLVM 83  
 Db 61 mkriqdcyvenngslisryldglvm 83

RESULT 17  
 AAY51471  
 ID AAY51471 standard; Protein; 109 AA.

XX AAY51471;

DT 22-MAY-2000 (first entry)

DE Human TRFP chain 2 (short form) protein fragment.

KM T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
 KM down regulation; immune response; allergen; immunoglobulin E;  
 KM sensitivity; cat protein allergen; human; chain 2.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (TMMD-) IMMUNOLOGIC PHARM CORP.

PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;

DR WPI: 2000-146862/13.

DR N-PSDB; AA288618.

XX Peptides of human T cell reactive feline protein for treating  
 XX sensitivity to cat protein allergens comprise at least one T cell  
 XX epitope recognized by a T cell receptor specific for the human T cell  
 XX reactive feline protein -  
 XX Claim 1; Column 79-80; 105bp; English.

CC This invention describes a novel peptide (I) of human T cell reactive  
 CC feline protein (hTRFP) having at least one T cell epitope recognized

CC by a T cell receptor specific for the human T cell reactive feline  
 CC protein, the peptide consisting of at least 7-30 amino acids, and having  
 CC an amino acid sequence derived from an amino acid sequence comprising 94,  
 CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
 CC down regulate the immune response to the allergen. The peptides have  
 CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
 CC peptide (I) is useful in compositions for treating sensitivity to a cat  
 CC protein allergen in a subject. This sequence represents the human TRFP  
 CC chain 2 (short form).  
 XX  
 SO Sequence 109 AA;

Query Match 74.8%; Score 83; DB 21; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-73;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTMRGALLVALLVTOALGVKMAETCPFYDVFPAVANGNELLDLSITKVNATEPERTA 60  
 Db 1 dtmrgallvllalvtqalgvkaetcpfydvffavangnellldlsitkvnateperpta 60  
 QY 61 MKKIDCYVENGILSRVLDGLVM 83  
 Db 61 mkkidcyvengilsrvidglvm 83

## RESULT 18

AA27370  
 ID AAR27370 standard; protein; 110 AA.

AC AAR27370;

DF 20-MAY-1998 (first entry)

DE TRFP Chain #2 with C2 leader sequence.

KW T cell reactive feline protein; cat allergy; allergic; IgE;  
 KM desensitizing.

OS Felis domesticus.

FH Key Location/Qualifiers

FT Peptide 1..19 /label- C2 leader

FT Protein 19..110 /label- TRFP chain #2

XX WO9215613-A.

PD 17-SEP-1992.

PF 20-FEB-1992; 92WO-US01344.

PR 28-FEB-1991; 91US-0662193.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Bond J, Kuo M;

DR WPI: 1992-331670/40.

PT Modified human T-cell reactive feline protein - stimulates T-cell  
 PT in individuals allergic to cats and shows reduced  
 PT histamine-releasing properties  
 PS Claim 1; Fig 1; 35pp; English.

CC This sequence represents a modified human T-cell reactive feline  
 CC protein which stimulates T-cells from an individual who is allergic  
 CC to cats, but which interacts with human IgE to a lesser extent than  
 CC does affinity purified TRFP. The protein is modified by treating  
 CC with either a mild alkali (pH 12.5-13.5 KOH, NaOH, LiOH or tetrabutyl  
 CC amines) or an enzyme which removes O-linked groups (cardiolipate

CC moieties). It is useful in desensitizing people who are allergic to cats.  
 XX  
 SO Sequence 110 AA;

Query Match 74.8%; Score 83; DB 13; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-73;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTMRGALLVALLVTOALGVKMAETCPFYDVFPAVANGNELLDLSITKVNATEPERTA 60  
 Db 1 dtmrgallvllalvtqalgvkaetcpfydvffavangnellldlsitkvnateperpta 60  
 QY 61 MKKIDCYVENGILSRVLDGLVM 83  
 Db 61 mkkidcyvengilsrvidglvm 83

## RESULT 19

AA28935  
 ID AAB28935 standard; Protein; 97 AA.

AC AAB28935;

DT 29-JAN-2001 (first entry)

DE T cell reactive feline protein chain 2 truncated form.

KW Cat; allergy; human T cell reactive feline protein; hTRFP;  
 KM Immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 95US-0431184.

PR 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Gefter ML, Garman RD, Greenstein JL, Bond JF;

DR WPI: 2000-601477/57.

DR N-PSDB: AAC60105.

PT Detecting, preventing and treating sensitivity to cat protein allergen  
 PT comprises combining a biological sample with a human T cell reactive  
 PT feline protein and determining the extent of binding that occurs -

PS Claim 1; Figure 5; 106pp; English.

CC The present invention relates to the detection of sensitivity to a cat  
 CC protein allergen by combining a blood sample from a subject with a  
 CC peptide of human T cell reactive feline protein (hTRFP). This method  
 CC and the hTRFP peptides are useful for diagnosing, preventing and  
 CC treating cat allergies by reducing or abolishing an individual's  
 CC allergic response to a cat allergen. DNA encoding the TRFP may be  
 CC used as probes to locate equivalent sequences present in other species.  
 CC These may further be used to study the mechanism of immunotherapy of  
 CC cat allergy, and to design modified derivatives, analogues or  
 CC functional equivalents useful in immunotherapy. The present  
 CC sequence was used in the invention.

XX Sequence 97 AA;

Query Match 73.0%; Score 81; DB 21; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-71;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTMRGALLVTLALVTQALGVKMAETCPFYDVFFAVANGNELLDLSITKVNATEPERETA 60  
 Db 1 dtmrgallvltallvtqalgvkmaetcpfydvffavangnellldlstkvnatepereta 60

OY 61 MKRIQDCYVENGSLISRYLDGL 81  
 Db 61 mkriqdcyvenslisyldgl 81

RESULT 20  
 AAY87675  
 ID AAY87675 standard; Protein: 97 AA.  
 AC AAY87675;  
 XX  
 DT 22-AUG-2000 (first entry)  
 XX  
 DE Feline human TRFP chain 2 truncated form protein.  
 XX  
 KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
 KM antiallergic; T cell stimulator; diagnostic; immunotherapy.  
 XX  
 OS Felis sp.  
 XX  
 PN US6048962-A.  
 XX  
 PD 11-Apr-2000.  
 XX  
 PE 27-APR-1995; 95US-0430014.  
 XX  
 PR 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Kuo M, Rogers BL, Gefter ML, Morgenstern JP, Brauer AW;  
 PI Greenstein JL, Griffith IJ, Garman RD;  
 XX  
 DR N-PSDB; AAA12246.  
 XX  
 PT New human T cell reactive feline protein useful for reducing or  
 PT abolishing individual's allergic response to cat allergen comprising  
 PT two different covalently linked peptide chains -  
 XX  
 PS Claim 2; Column 79-80; 106pp; English.  
 XX  
 CC This invention describes a novel naturally occurring cat protein allergen  
 CC (1), human T cell reactive feline protein (TRFP), comprising two  
 CC different covalently linked peptide chains with a molecular weight of 20  
 CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
 CC under reducing conditions. The products of the invention have  
 CC antiallergic activity and act as human T cell stimulators. TRFP is useful  
 CC for reducing or preventing the adverse effects of cat allergens on cat  
 CC allergic individuals and in ex vivo diagnostic tests to determine which  
 CC peptides cause sensitivity so as to selectively use them to desensitize  
 CC a cat sensitive individual. Purified TRFP is also useful for studying  
 CC the mechanism of immunotherapy of cat allergy and to design modified  
 CC derivatives, analogs or functional equivalents that are more useful in  
 CC immunotherapy against cat allergy. DNA sequences encoding TRFP are  
 CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics

CC and/or therapeutics. Fully defined and characterized TRFP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d I)  
 CC chain 2, truncated form which is described in the method of the  
 CC invention.  
 CC  
 SQ Sequence 97 AA;  
 Query Match 73.0%; Score 81; DB 21; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-71;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTMRGALLVTLALVTQALGVKMAETCPFYDVFFAVANGNELLDLSITKVNATEPERETA 60  
 Db 1 dtmrgallvltallvtqalgvkmaetcpfydvffavangnellldlstkvnatepereta 60

OY 61 MKRIQDCYVENGSLISRYLDGL 81  
 Db 61 mkriqdcyvenslisyldgl 81

RESULT 21  
 AAY90105  
 ID AAY90105 standard; Protein: 97 AA.  
 AC AAY90105;  
 XX  
 DT 13-JUL-2000 (first entry)  
 XX  
 DE Cat TRFP chain 2 truncated form protein sequence.  
 XX  
 KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;  
 KM house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;  
 KW diagnosis; goat; sheep; horse; rabbit; dog.  
 XX  
 OS Felis domesticus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..19  
 FT /note= "signal peptide"  
 FT Protein 20..97  
 FT /note= "mature TRFP chain 2 truncated form"  
 XX  
 PN US6025162-A.  
 XX  
 PD 15-FEB-2000.  
 XX  
 PE 28-APR-1995; 95US-0430944.  
 XX  
 PR 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Morgenstern JP, Griffith IJ, Rogers BL;  
 PI Greenstein JL, Griffith IJ, Garman RD;  
 XX  
 DR N-PSDB; AAA07439.  
 XX  
 PT New human T cell reactive feline protein, useful for desensitizing cat  
 PT allergic individuals to cat allergens -  
 XX  
 PS Claim 1; Fig 5; 108pp; English.  
 XX  
 CC This sequence is a peptide chain of the human T cell reactive feline  
 CC protein (TRFP) of the invention. The protein is a cat protein allergen,  
 CC and was isolated from a vacuum bag extract obtained by affinity





CC can be used to reduce/eliminate the allergic response partic. by  
 CC modifcn. of lymphokine prodn. by the T-cells. They can also be  
 CC used to identify epitopes responsible for sensitivity. The DNA can  
 CC be used to detect comparable sequence in other species, and also  
 CC for prodn. of modified forms of TRFP esp. showing reduced binding  
 CC to IgE and thus reduced tendency to cause adverse reactions.  
 CC See also AAI2119-R12123.

SO Sequence 109 AA;

Query Match 71.2%; Score 79; DB 12; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-69;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DTRGALIVALLVTOALGKMAETCFIFDVFPFPAVANGNELLDLSLTVNATEPERTA 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 1 dtrmgalllvallvtqalgvmaetcpifdvifavangnellldsltkvnateperta 60

OY 61 MKRIODCYENGILSRVLD 79  
 ||||||||||||||||||||  
 DB 61 mkridcyengilstrvld 79

RESULT 24

AAW27382  
 ID AAW27382 standard; protein; 92 AA.

AAW27382;

09-APR-1998 (first entry)

Chain 2 of major cat allergen Fel d1.

Major cat allergen; Fel d1; chain 2; cryptic peptide; T-cell; asthma;  
 atopic allergy; therapy.

Felis sp.

MO9735193-A1.

25-SEP-1997.

20-MAR-1997; 97MO-GB00783.

24-APR-1996; 96GB-0008430.

21-MAR-1996; 96GB-0005904.

(UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.

Kay AB, Larche M;

WPI; 1997-480354/44.

Determining if peptide of protein is cryptic peptide - by comparing  
 its reactivity with pre-challenged and non-pre-challenged T cells,  
 useful to diagnose or treat atopic condition, e.g. asthma

Claim 12; Page 19; 49pp; English.

This sequence represents the chain 2 sequence of the major cat allergen  
 Fel d1. This sequence can be used in the method of the invention. The  
 method of the invention is for determining if a peptide of a protein is a  
 cryptic peptide, and comprises: (a) exposing T-cells to the peptide in a  
 primary challenge, and measuring the reactivity of the T-cells to the  
 peptide; (b) exposing pre-challenged T-cells, obtained by exposure to the  
 protein, to the peptide in a secondary challenge, and measuring the  
 reactivity of the pre-challenged T-cells to the peptide; and  
 (c) determining the peptide to be a cryptic peptide if T-cell reactivity  
 is observed in the secondary, but not in the primary challenge. Peptides  
 identified as being cryptic can be used to diagnose or treat an atopic  
 allergy, e.g. asthma. They can also be used in to screen compounds for  
 therapeutic activity, e.g. asthma therapeutic activity.

XX Sequence 92 AA;

Query Match 59.5%; Score 66; DB 18; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-56;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 LSLTKVNATEPERTAMKRIODCYENGILSRVLDGLVMTISSSKDQMGFAVONTVEDLK 105  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 27 lsltkvnatepertamkridcyengilstrvldglvmtlssskdcmgeavqntvedlk 86

OY 106 LNTLGR 111  
 |||||||  
 DB 87 lntlgr 92

RESULT 25

AAV25532  
 ID AAV25532 standard; peptide; 92 AA.

AAV25532;

30-SEP-1999 (first entry)

Human MHC Class II desensitizing peptide #6.

Major histocompatibility complex; class II; desensitizing; human;  
 allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
 chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
 screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
 cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
 mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

Synthetic.

MO9934826-A1.

15-JUL-1999.

11-JAN-1999; 99MO-GB00080.

21-SEP-1998; 98GB-0020474.

09-JAN-1998; 98GB-0000445.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Kay AB, Larche M;

WPI; 1999-458255/38.

Desensitizing patients to polypeptide allergens

Claim 7; Figure 9; 117pp; English.

This invention describes a novel method of desensitizing a patient to a  
 polypeptide allergen and comprises administering to the patient a peptide  
 derived from the allergen where restriction to a MHC Class II molecule  
 possessed by the patient can be demonstrated for the peptide and the  
 peptide is able to induce a late phase response in an individual who  
 possesses the MHC Class II molecule. The methods can be used for  
 desensitizing patients to allergens present in e.g. grass, tree and weed  
 (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
 the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
 fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
 non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
 the annelid mollusc earthworm, mammals such as cat, dog, horse, cow, pig,  
 rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
 produce immunological vaccines which may be used to prevent and/or treat  
 conditions involving hypersensitivity to allergens. This sequence  
 represents a peptide used to desensitise the human major  
 histocompatibility complex (MHC) class II response to the Fel d1 Chain 2  
 allergen.

XX Sequence 92 AA:  
 SO Query Match 59.5%; Score 66; DB 20; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-56;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 46 LSLTVNATEPPTAMKTIQDCYVENGLISVTDGLVMTTSSSKDCMGEAVONTVEDLK105  
 DB 27 LSLTVNATEPPTAMKTIQDCYVENGLISVTDGLVMTTSSSKDCMGEAVONTVEDLK105  
 OY 106 LMTLGR 111  
 DB 87 Intlgtr 92  
 RESULT 26  
 AAB28938  
 ID AAB28938 standard; Protein: 90 AA.  
 AC AAB28938;  
 XX  
 DT 29-JAN-2001 (first entry)  
 XX  
 DE T cell reactive feline protein chain 2 PRO short.  
 XX  
 KW Cat; allergy: human T cell reactive feline protein; hTRFP;  
 immunotherapy.  
 XX  
 OS Fells sp.  
 XX  
 PN US6120769-A.  
 PD 19-SEP-2000.  
 XX  
 PF 28-APR-1995; 950S-0431184.  
 XX  
 PR 02-SEP-1994; 940S-0300928.  
 PR 03-NOV-1989; 890S-0431565.  
 PR 28-FEB-1991; 910S-0662276.  
 PR 13-DEC-1991; 910S-0807529.  
 PR 25-MAR-1992; 920S-0857311.  
 PR 15-MAY-1992; 920S-0884718.  
 PR 15-JAN-1993; 930S-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Geffer ML, Garman RD, Greenstein JL, Bond JF;  
 DR WPI: 2000-601477/57.  
 PT Detecting, preventing and treating sensitivity to cat protein allergen  
 PT comprises combining a biological sample with a human T cell reactive  
 PT feline protein and determining the extent of binding that occurs  
 XX  
 PS Disclosure; Figure 7; 106pp; English.  
 XX  
 CC The present invention relates to the detection of sensitivity to a cat  
 CC protein allergen by combining a blood sample from a subject with a  
 CC peptide of human T cell reactive feline protein (hTRFP). This method  
 CC and the hTRFP peptides are useful for diagnosing, preventing and  
 CC treating cat allergies by reducing or abolishing an individual's  
 CC allergic response to a cat allergen. DNA encoding the TRFP may be  
 CC used as probes to locate equivalent sequences present in other species.  
 CC These may further be used to study the mechanism of immunotherapy of  
 CC cat allergy, and to design modified derivatives, analogues or  
 CC functional equivalents useful in immunotherapy. The present  
 CC sequence was used in the invention.  
 CC  
 CC Sequence 90 AA:

XX Query Match 57.7%; Score 64; DB 21; Length 90;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
 XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 20 VKNAETCPFYVFPVAVANGNELLDLSLTXYNATEPPTAMKTIQDCYVENGLISVTD 79  
 DB 1 VKNAETCPFYVFPVAVANGNELLDLSLTXYNATEPPTAMKTIQDCYVENGLISVTD 60  
 OY 80 GLVW 83  
 DB 61 glvwm 64  
 RESULT 27  
 AAY87679  
 ID AAY87679 standard; Protein: 90 AA.  
 AC AAY87679;  
 XX  
 DT 22-AUG-2000 (first entry)  
 XX  
 DE Feline human TRFP chain 2 short form protein #2.  
 XX  
 KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
 KW antiallergic; T cell stimulator; diagnostic; immunotherapy.  
 XX  
 OS Fells sp.  
 XX  
 PN US6048962-A.  
 PD 11-APR-2000.  
 XX  
 PF 27-APR-1995; 950S-0430014.  
 XX  
 PR 02-SEP-1994; 940S-0300928.  
 PR 03-NOV-1989; 890S-0431565.  
 PR 28-FEB-1991; 910S-0662276.  
 PR 13-DEC-1991; 910S-0807529.  
 PR 25-MAR-1992; 920S-0857311.  
 PR 15-MAY-1992; 920S-0884718.  
 PR 15-JAN-1993; 930S-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;  
 PI Greenstein JL, Griffith ID, Garman RD;  
 DR WPI: 2000-316905/27.  
 PT New human T cell reactive feline protein useful for reducing or  
 PT abolishing individual's allergic response to cat allergen comprising  
 PT two different covalently linked peptide chains  
 XX  
 PS Example 2; Column 83-84; 106pp; English.  
 XX  
 CC This invention describes a novel naturally occurring cat protein allergen  
 CC (I), human T cell reactive feline protein (TRFP), comprising two  
 CC different covalently linked peptide chains with a molecular weight of 20  
 CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
 CC under reducing conditions. The products of the invention have  
 CC antiallergic activity and act as human T cell stimulators. TRFP is useful  
 CC for reducing or preventing the adverse effects of cat allergens on cat  
 CC allergic individuals and in ex vivo diagnostic tests to determine which  
 CC peptides cause sensitivity so as to selectively use them to desensitize  
 CC a cat sensitive individual. Purified TRFP is also useful for studying  
 CC the mechanism of immunotherapy of cat allergy and to design modified  
 CC derivatives, analogs or functional equivalents that are more useful in  
 CC immunotherapy against cat allergy. DNA sequences encoding TRFP are  
 CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
 CC and/or therapeutics. Fully defined and characterized TRFP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d I)

CC chain 2, short form which is described in the method of the invention.  
XX  
SQ Sequence 90 AA;

Query Match 57.7%; Score 64; DB 21; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 20 VKNAETCPIDYDFEAVANGNELLDLSLRKVATPEPTAMKIKDCYENGILISVLD 79  
Db 1 vknaetcpifdydfavangnellldsltkvnatepertamkikdcyengilistrld 60  
OY 80 GLVM 83  
Db 61 glvm 64

## RESULT 28

AAVS1476  
ID AAVS1476 standard; Protein; 90 AA.

AC AAVS1476;

DT 22-MAY-2000 (first entry)

DE Human TRFP chain 2 (short form) protein fragment #2.

XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
KM down regulation; immune response; allergen; immunoglobulin E;  
KW sensitivity; cat protein allergen; human; chain 2.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PE 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;

XX WPI; 2000-146862/13.

XX Peptides of human T cell reactive feline protein for treating

XX PT sensitivity to cat protein allergens comprise at least one T cell

XX PT epitope recognized by a T cell receptor specific for the human T cell

XX PT reactive feline protein -

XX Example 1; Column 83-84; 105pp; English.

XX This invention describes a novel peptide (I) of human T cell reactive

XX CC feline protein (hTRFP) having at least one T cell epitope recognized

XX CC by a T cell receptor specific for the human T cell reactive feline

XX CC protein, the peptide consisting of at least 7-30 amino acids, and having

XX CC an amino acid sequence derived from an amino acid sequence comprising 94,

XX CC 96, 97, 109, or 111 residues, given in the specification. The peptides

XX CC down regulate the immune response to the allergen. The peptides have

XX CC reduced immunoglobulin E binding and reduce T cell responsiveness. The

XX CC peptide (I) is useful in compositions for treating sensitivity to a cat

XX CC protein allergen in a subject. This sequence represents the human TRFP

XX CC chain 2 (short form).

XX Sequence 90 AA;

Query Match 57.7%; Score 64; DB 21; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 VKNAETCPIDYDFEAVANGNELLDLSLRKVATPEPTAMKIKDCYENGILISVLD 79  
Db 1 vknaetcpifdydfavangnellldsltkvnatepertamkikdcyengilistrld 60  
OY 80 GLVM 83  
Db 61 glvm 64

## RESULT 29

AAAR27372  
ID AAR27372 standard; protein; 101 AA.

AC AAR27372;

DT 20-MAY-1998 (first entry)

DE TRFP Chain #2 C2ST with C2 leader sequence.

KW T cell reactive feline protein; cat allergy; allergic; IgE;  
desensitizing.

OS Felis domesticus.

XX Key location/Qualifiers

FT peptide 1..19 /label= C2 leader

FT protein 20..101 /label= TRFP chain 2 PRO

FT misc\_difference 52 /note= "X not defined in specification"

FT misc\_difference 84 /label= Thr, Ile

FT misc\_difference 85 /label= Thr, Ala

FT misc\_difference 87 /label= Ser, Asn

FT misc\_difference 88 /label= Ser, Glu

FT misc\_difference 99 /label= Thr, Ala

FT misc\_difference 100 /label= Val, Met

XX WO9215613-A.

XX 17-SEP-1992.

XX 20-FEB-1992; 92WO-US01344.

XX 28-FEB-1991; 91US-0662193.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Bond J, Kuo M;

XX WPI; 1992-331670/40.

XX Modified human T-cell reactive feline protein - stimulates T-cell

XX PT in individuals allergic to cats and shows reduced

XX PT histamine-releasing properties

XX Claim 1; Fig 1; 35pp; English.

XX This sequence represents a modified human T-cell reactive feline

XX CC protein which stimulates T-cells from an individual who is allergic

XX CC to cats, but which interacts with human IgE to a lesser extent than



Query Match 33.3%; Score 37; DB 21; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SLTKVNATEPERTAMKKIDQCYVENGILSRVLDGLVM 83  
 DB 1 sltkvnatepertamkkidqcyvengilsrvldglvm 37

## RESULT 32

AA51477  
 ID AAY51477 standard; Protein; 51 AA.

AC AAY51477;

DT 22-MAY-2000 (first entry)

DE Human TRFP chain 2 (truncated form) protein fragment #2.

KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
 down regulation; immune response; allergen; immunoglobulin E;

KM sensitivity; cat protein allergen; human; chain 2.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Garman RD, Greenstein JL, Kuo M, Briner TU, Morville M, Gefter ML;

DR WPI; 2000-146862/13.

PT Peptides of human T cell reactive feline protein for treating

PT sensitivity to cat protein allergens comprise at least one T cell

PT epitope recognized by a T cell receptor specific for the human T cell

PT reactive feline protein -

PS Example 1; Column 85-86; 105pp; English.

XX This invention describes a novel peptide (I) of human T cell reactive

CC feline protein (hTRFP) having at least one T cell epitope recognized

CC by a T cell receptor specific for the human T cell reactive feline

CC protein, the peptide consisting of at least 7-30 amino acids, and having

CC an amino acid sequence derived from an amino acid sequence comprising 94,

CC 96, 97, 109, or 111 residues, given in the specification. The peptides

CC down regulate the immune response to the allergen. The peptides have

CC reduced immunoglobulin E binding and reduce T cell responsiveness. The

CC peptide (I) is useful in compositions for treating sensitivity to a cat

CC protein allergen in a subject. This sequence represents the human TRFP

CC chain 2 (truncated form).

XX Sequence 51 AA;

SQ

Query Match 33.3%; Score 37; DB 21; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SLTKVNATEPERTAMKKIDQCYVENGILSRVLDGLVM 83

DB 1 sltkvnatepertamkkidqcyvengilsrvldglvm 37

RESULT 33  
 ID AAR27371 standard; Protein; 82 AA.

AC AAR27371;

DT 20-MAY-1998 (first entry).

DE TRFP Chain #2 C2ST with C2 leader sequence.

KW T cell reactive feline protein; cat allergy; allergic; IGE;

KM desensitizing.

OS Felis domesticus.

FH Key Location/Qualifiers

FT peptide 1..19 /label= C2 leader

FT protein 20..82 /label= TRFP chain #2

PN WO9215613-A.

PD 17-SEP-1992.

PF 20-FEB-1992; 92WO-US01344.

PR 28-FEB-1991; 91US-0662193.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Bond J, Kuo M;

DR WPI; 1992-331670/40.

PT Modified human T-cell reactive feline protein - stimulates T-cell

PT in individuals allergic to cats and shows reduced

PT histamine-releasing properties

PS Claim 1; Fig 1; 35pp; English.

XX This sequence represents a modified human T-cell reactive feline

CC protein which stimulates T-cells from an individual who is allergic

CC to cats, but which interacts with human IgE to a lesser extent than

CC does affinity purified TRFP. The protein is modified by treating

CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary

CC amines) or an enzyme which removes O-linked groups (carbohydrate

CC moieties). It is useful in desensitizing people who are allergic to cats.

XX Sequence 82 AA;

SQ

Query Match 33.3%; Score 37; DB 13; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SLTKVNATEPERTAMKKIDQCYVENGILSRVLDGLVM 83  
 DB 20 sltkvnatepertamkkidqcyvengilsrvldglvm 56

## RESULT 34

AA28940  
 ID AAB28940 standard; Protein; 82 AA.

AC AAB28940;

DT 29-JAN-2001 (first entry)

XX



```
XX SQ Sequence 82 AA;
Query Match 28.8%; Score 32; DB 21; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.8e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 20 VKMAETCPITYDVFFAVANGNELLDLSLTKV 51
DB 1 vkmaetcpitydvffavangnellldlsltk 32

RESULT 36
AA90106
ID AA90106 standard; Protein; 82 AA.
AC AA90106;
XX
DT 13-JUL-2000 (first entry)
XX
DE Cat TRFP chain 2 generic protein sequence.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KM house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
XX diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
FH Key Location/Qualifiers
FT MISC-difference 33
FT MISC-difference 65 /note="unspecified amino acid"
FT MISC-difference 65 /label= Ile, Thr
FT MISC-difference 66 /label= Ala, Thr
FT MISC-difference 68 /label= Asn, Ser
FT MISC-difference 69 /label= Glu, Ser
FT MISC-difference 80 /label= Ala, Thr
FT MISC-difference 81 /label= Met, Val
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PF 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Morgenstern JP, Griffith IU, Rogers BL;
DR WPI; 2000-181812/16.
PT New human T cell reactive feline protein, useful for desensitizing cat
XX allergic individuals to cat allergens -
XX
PS Claim 1; Column 85-86; 108pp; English.
XX
CC This sequence is a peptide chain of the human T cell reactive feline
CC protein (TRFP) of the invention. The protein is a cat protein allergen,
CC and was isolated from a vacuum bag extract obtained by affinity
```

```
CC purification of house dust collected from several homes with cats. TRFP
CC is composed of two covalently linked peptide chains, and is also referred
CC to as Fel d I. TRFP and its peptides are useful for reducing or
CC preventing the adverse effects that exposure to cat allergens normally
CC has on cat allergic individuals (i.e. to desensitize individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in a diagnostic and/or therapeutic applications.
XX
SQ Sequence 82 AA;
Query Match 28.8%; Score 32; DB 21; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.8e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 20 VKMAETCPITYDVFFAVANGNELLDLSLTKV 51
DB 1 vkmaetcpitydvffavangnellldlsltk 32

RESULT 37
AA51478
ID AA51478 standard; Protein; 82 AA.
XX
AC AA51478;
XX
DT 22-MAY-2000 (first entry)
XX
DE Human TRFP chain 2 protein fragment.
XX
FH Key Location/Qualifiers
FT MISC-difference 33
FT MISC-difference 65 /label= unidentified
FT MISC-difference 66 /label= Ile, Thr
FT MISC-difference 66 /label= Ala, Thr
FT MISC-difference 68 /label= Asn, Ser
FT MISC-difference 69 /label= Glu, Ser
FT MISC-difference 80 /label= Ala, Thr
FT MISC-difference 81 /label= Met, Val
XX
PN US6019972-A.
XX
PD 01-FEB-2000.
XX
PF 02-SEP-1994; 94US-0300928.
XX
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Garman RD, Greenstein JL, Kuo M, Briner TU, Morville M, Gelfer ML;
XX WPI; 2000-146862/13.
DR
```

```
XX Peptides of human T cell reactive feline protein for treating
PT sensitivity to cat protein allergens comprise at least one T cell
PT epitope recognized by a T cell receptor specific for the human T cell
PT reactive feline protein
XX
XX Example 1; Column 85-86; 105bp; English.
XX
CC This invention describes a novel peptide (1) of human T cell reactive
CC feline protein (hrfpp) having at least one T cell epitope recognized
CC by a T cell receptor specific for the human T cell reactive feline
CC protein, the peptide consisting of at least 7-30 amino acids, and having
CC an amino acid sequence derived from an amino acid sequence comprising 94,
CC 96, 97, 109, or 111 residues given in the specification. The peptides
CC down regulate the immune response to the allergen. The peptides have
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The
CC peptide (1) is useful in compositions for treating sensitivity to a cat
CC protein allergen in a subject. This sequence represents the human TRFP
CC chain 2.
XX
XX Sequence 82 AA:
SQ
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 32; DB 21; Length 82;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 20 VKMAETCPFYDVFVAVANGNELLDLSITRVV 51
XX ||||||||||||||||||||||||||||
XX 1 VKMAETCPFYDVFVAVANGNELLDLSITKV 32
XX
XX RESULT 38
XX AAR36544
XX ID AAR36544 standard; Protein; 26 AA.
XX
XX AAR36544:
XX
XX 12-AUG-1993 (first entry)
XX
XX Peptide 2.
XX
XX Human T cell reactive feline protein; TRFP; epitope; recombinant.
XX
XX Fells.
XX
XX WO9308280-A.
XX
XX 29-APR-1993.
XX
XX 16-OCT-1992; 92WO-US08694.
XX
XX 16-OCT-1991; 91US-077859.
XX
XX 13-DEC-1991; 91US-0807529.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
XX Rogers BL;
XX WPI; 1993-152473/18.
XX
XX Recombinant peptide having T-cell stimulating activity - for the
XX diagnosis and treatment of sensitivity to protein allergens,
XX autoantigens and protein antigens
XX
XX Disclosure; Fig 4; 73pp; English.
XX
XX Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
XX coli and purified. T cell epitope studies using overlapping peptide
XX regions derived from the TRFP amino acids sequence were used to
XX identify multiple T cell epitopes in each chain of TRFP. DNA
XX constructs were assembled in which 3 regions (encoding peptides X,
```

```
CC Y and Z) were linked to produce DNA constructs encoding recombinant-
CC peptides.
XX
XX Sequence 26 AA:
SQ
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 26; DB 14; Length 26;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 33 FFVANGNELLDLSITRVVATPEPR 58
XX ||||||||||||||||||||||||
XX 1 FFVANGNELLDLSITKVATPEPR 26
XX
XX RESULT 39
XX AAR41977
XX ID AAR41977 standard; peptide; 26 AA.
XX
XX AAR41977:
XX
XX 21-APR-1994 (first entry)
XX
XX Human T cell reactive feline protein fragment 2.
XX
XX Human; T cell; reactive; feline; protein; immune response; antigen;
XX tolerance; mammal; Dermatophagoides; Fells; Ambrosia; Lolium; Canis;
XX Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemisia;
XX Plantago; Parietaria; Blatella; Apsis; Periphaneta; autoantigen; ss.
XX
XX Homo sapiens.
XX
XX WO9319178-A.
XX
XX 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US02462.
XX
XX 25-MAR-1992; 92US-0857311.
XX
XX 15-MAY-1992; 92US-0884718.
XX
XX 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Briner TJ, Garman RD, Gelfer ML, Greenstein JL;
XX Kuo M, Morville M;
XX WPI; 1993-320744/40.
XX
XX New peptide(s) for inducing tolerance - comprise one or more
XX epitope(s) of an allergen administered subcutaneously, for
XX treating sensitivity to cats, bees, etc.
XX
XX Claim 1; Fig 3; 107pp; English.
XX
XX The sequences given in AAR41975-82 are peptides derived from a human T
XX cell reactive feline protein. These peptides are used in a
XX therapeutic composition which is useful in treating diseases which
XX involve an immune response to a protein antigen. This composition
XX may be used to induce tolerance in a mammal to Dermatophagoides,
XX Fells, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
XX Quercus, Olea, Artemisia, Plantago, Parietaria, Canis, Blatella,
XX Apsis, Periphaneta and to autoantigens in humans.
XX
XX Sequence 26 AA:
SQ
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 26; DB 14; Length 26;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 33 FFVANGNELLDLSITRVVATPEPR 58
XX ||||||||||||||||||||||||
```



Db 1 ffavangnell1d1sltkvateper 26

RESULT 40

ID AAB28943 standard; Peptide; 26 AA.

XX AAB28943;

XX AAB28943;

XX 29-JAN-2001 (first entry)

XX Peptide 2 derived from T cell reactive feline protein.

XX Cat; allergy; human T cell reactive feline protein; hTRFP;

XX Immunotherapy.

XX Felis sp.

XX US6120769-A.

XX 19-SEP-2000.

XX 28-APR-1995; 95US-0431184.

XX 02-SEP-1994; 94US-0300928.

XX 03-NOV-1989; 89US-0431565.

XX 28-FEB-1991; 91US-0662276.

XX 13-DEC-1991; 91US-0807529.

XX 25-MAR-1992; 92US-0857311.

XX 15-MAY-1992; 92US-0884718.

XX 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Geffer ML, Garman RD, Greenstein JL, Bond JF;

XX WPI; 2000-601477/57.

XX Detecting, preventing and treating sensitivity to cat protein allergen

XX comprises combining a biological sample with a human T cell reactive

XX feline protein and determining the extent of binding that occurs

XX Claim 2; Figure 17; 106pp; English.

XX The present invention relates to the detection of sensitivity to a cat

XX protein allergen by combining a blood sample from a subject with a

XX peptide of human T cell reactive feline protein (hTRFP). This method

XX and the hTRFP peptides are useful for diagnosing, preventing and

XX treating cat allergies by reducing or abolishing an individual's

XX allergic response to a cat allergen. DNA encoding the TRFP may be:

XX used as probes to locate equivalent sequences present in other species.

XX These may further be used to study the mechanism of immunotherapy of

XX cat allergy, and to design modified derivatives, analogues or

XX functional equivalents useful in immunotherapy. The present

XX sequence was used in the invention.

XX Sequence 26 AA;

XX SQ

Query Match 23.4%; Score 26; DB 21; Length 26;

Best Local Similarity 100.0%; Pred. No. 3.9e-18;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 FFAVANGNELL1D1SLTKVATEPER 58

Db 1 ffavangnell1d1sltkvateper 26

RESULT 41

AAV87684

ID AAV87684 standard; Protein; 26 AA.

XX AAV87684;

XX 22-AUG-2000 (first entry)

XX Feline human TRFP peptide 2.

XX T-cell reactive feline protein; TRFP; Fel d I; cat allergen;

XX antiallergic; T cell stimulator; diagnostic; immunotherapy.

XX Felis sp.

XX US6048962-A.

XX 11-APR-2000.

XX 27-APR-1995; 95US-0430014.

XX 02-SEP-1994; 94US-0300928.

XX 03-NOV-1989; 89US-0431565.

XX 28-FEB-1991; 91US-0662276.

XX 13-DEC-1991; 91US-0807529.

XX 25-MAR-1992; 92US-0857311.

XX 15-MAY-1992; 92US-0884718.

XX 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;

XX Greenstein JL, Griffith JI, Garman RD;

XX WPI; 2000-316905/27.

XX New human T cell reactive feline protein useful for reducing or

XX abolishing individual's allergic response to cat allergen comprising

XX two different covalently linked peptide chains

XX Example 5; Column 87-88; 106pp; English.

XX This invention describes a novel naturally occurring cat protein allergen

XX (I), human T cell reactive feline protein (TRFP), comprising two

XX different covalently linked peptide chains with a molecular weight of 20

XX kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD

XX under reducing conditions. The products of the invention have

XX antiallergic activity and act as human T cell stimulators. TRFP is useful

XX for reducing or preventing the adverse effects of cat allergens on cat

XX allergic individuals and in ex vivo diagnostic tests to determine which

XX peptides cause sensitivity so as to selectively use them to desensitize

XX a cat sensitive individual. Purified TRFP is also useful for studying

XX the mechanism of immunotherapy of cat allergy and to design modified

XX derivatives, analogs or functional equivalents that are more useful in

XX immunotherapy against cat allergy. DNA sequences encoding TRFP are

XX useful as probes to locate equivalent sequences present in other species

XX (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics

XX and/or therapeutics. Fully defined and characterized TRFP provides

XX complete and a very simple desensitization therapy. This sequence

XX represents a human T cell reactive feline protein (also known as Fel d I)

XX derived peptide 2 which is described in the method of the invention.

XX Sequence 26 AA;

XX SQ

Query Match 23.4%; Score 26; DB 21; Length 26;

Best Local Similarity 100.0%; Pred. No. 3.9e-18;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 FFAVANGNELL1D1SLTKVATEPER 58

Db 1 ffavangnell1d1sltkvateper 26

RESULT 42

AAV90109

ID AAV90109 standard; Peptide; 26 AA.

XX AAV90109;

AC AAY90109;  
 XX 13-JUL-2000 (first entry)  
 DT XX  
 DE Cat TRFP derived peptide, peptide 2.  
 XX  
 XX Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;  
 KW house dust; Fel d 1; cat allergy; Felis domesticus sensitivity; therapy;  
 KW diagnosis; goat; sheep; horse; rabbit; dog.  
 XX  
 OS Felis domesticus.  
 PN US6025162-A.  
 PD 15-FEB-2000.  
 XX  
 XX 28-APR-1995; 95US-0430944.  
 XX  
 XX 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX Morgenstern JP, Griffiths IJ, Rogers BL;  
 PI WPI: 2000-181812/16.  
 XX  
 PT New human T cell reactive feline protein, useful for desensitizing cat  
 PT allergic individuals to cat allergens -  
 XX  
 PS Claim 9; Fig 17; 108pp; English.  
 CC This sequence is a peptide derived from the human T cell reactive feline  
 CC protein (TRFP) of the invention. The protein is a cat protein allergen,  
 CC and was isolated from a vacuum bag extract obtained by affinity  
 CC purification of house dust collected from several homes with cats. TRFP  
 CC is composed of two covalently linked peptide chains, and is also referred  
 CC to as Fel d 1. TRFP and its peptides are useful for reducing or  
 CC preventing the adverse effects that exposure to cat allergens normally  
 CC has on cat allergic individuals (i.e. to desensitize individuals to cat  
 CC allergens or block the effect of the allergens). TRFP is also used in  
 CC methods of diagnosing sensitivity to feline domesticus in an individual.  
 CC DNA sequences encoding TRFP can be used as probes to locate equivalent  
 CC sequences present in other species, e.g. goat, sheep, horse, rabbit and  
 CC dog, that may be useful in diagnostic and/or therapeutic applications.  
 XX  
 SQ Sequence 26 AA;  
 Query Match 23.4%; Score 26; DB 21; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-18;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 33 FFVANGNELLDLSITKVNATEPER 58  
 DB 1 ffavangnellldlsitkvnateper 26  
 RESULT 43  
 AAY51481;  
 ID AAY51481 standard; Protein; 26 AA.  
 XX  
 XX AAY51481;  
 DT 22-MAY-2000 (first entry)  
 XX  
 XX Human TRFP derived peptide 2.  
 DE

KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
 KW down regulation; immune response; allergen; immunoglobulin E;  
 KW sensitivity; cat protein allergen; human.  
 XX  
 OS Homo sapiens.  
 PN US6019972-A.  
 PD 01-FEB-2000.  
 XX  
 XX 02-SEP-1994; 94US-0300928.  
 XX  
 XX 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Getter ML;  
 PI WPI: 2000-146862/13.  
 XX  
 PT Peptides of human T cell reactive feline protein for treating  
 PT sensitivity to cat protein allergens comprise at least one T cell  
 PT epitope recognized by a T cell receptor specific for the human T cell  
 PT reactive feline protein -  
 XX  
 PS Claim 8; Column 89-90; 105pp; English.  
 CC This invention describes a novel peptide (I) of human T cell reactive  
 CC feline protein (hTRFP) having at least one T cell epitope recognized  
 CC by a T cell receptor specific for the human T cell reactive feline  
 CC protein. The peptide consisting of at least 7-30 amino acids, and having  
 CC an amino acid sequence derived from an amino acid sequence comprising 94,  
 CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
 CC down regulate the immune response to the allergen. The peptides have  
 CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
 CC peptide (I) is useful in compositions for treating sensitivity to a cat  
 CC protein allergen in a subject. This sequence represents the human TRFP  
 CC derived peptide used in the method of the invention.  
 XX  
 SQ Sequence 26 AA;  
 Query Match 23.4%; Score 26; DB 21; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-18;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 33 FFVANGNELLDLSITKVNATEPER 58  
 DB 1 ffavangnellldlsitkvnateper 26  
 RESULT 44  
 AAR36548  
 ID AAR36548 standard; Protein; 96 AA.  
 XX  
 XX AAR36548;  
 DT 12-AUG-1993 (first entry)  
 XX  
 XX Recombitope yzx.  
 DE  
 XX Human T cell reactive feline protein; TRFP; epitope; recombitope  
 KW sensitivity; Felis domesticus.  
 XX  
 XX Synthetic.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Cleavage-site 14..15

```

FT      /label= thrombin_cleavage_site
XX
XX      WO9308280-A.
PN
XX
XX      29-APR-1993.
PD
XX
XX      16-OCT-1992; 92WO-US08694.
PF
XX      16-OCT-1991; 91US-0777859.
PR      13-DEC-1991; 91US-0807529.
PK
XX      (IMMU-) IMMUNOLOGIC PHARM CORP.
PA
XX      Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M.
PI      Rogers BL;
XX
XX      WPI; 1993-152473/18.
DR      N-PSDB; AA041572.
XX
XX      Recombitope peptide having T-cell stimulating activity - for the
PT      diagnosis and treatment of sensitivity to protein allergens,
PR      autoantigens and protein antigens
PS      Disclosure; Fig B; 73pp; English.
XX
XX      Preferred recombithope peptides for treating sensitivity to Felis
CC      domesticus are derived from the the genus Felis and comprise
CC      regions selected from peptides X, Y, Z, A and B, of TRFP, and
CC      modifications thereof, such as peptide C.
CC      Oligonucleotides C, D, E, F, G, H and I are used in the
CC      construction of recombithope peptide YZX.
XX
SQ      Sequence 96 AA:

Query Match      23.4%; Score 26; DB 14; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33 FFAVANGNELLDLSITKYNATEPER 58
      ||||||||||||||||||||||||||||
DB      44 ffavangnellldlsitkynateper 69

RESULT 45
ID      AAB28979 standard; Protein; 96 AA.
XX
XX      AAB28979;
AC
XX
XX      29-JAN-2001 (first entry)
DT
XX
XX      Peptide YZX.
DE
XX
XX      Cat; allergy; human T cell reactive feline protein; hTRFP;
KW      immunotherapy.
XX
XX      Fells sp.
OS      Synthetic.
OS
XX      US6120769-A.
PN
XX      19-SEP-2000.
PD
XX
XX      28-APR-1995; 95US-0431184.
PF
XX
XX      02-SEP-1994; 94US-0300928.
PR      03-NOV-1989; 89US-0431565.
PR      28-FEB-1991; 91US-0662276.
PR      13-DEC-1991; 91US-0807529.
PR      25-MAR-1992; 92US-0857311.
PR      15-MAY-1992; 92US-0884718.
PR      15-JAN-1993; 93US-0006116.

```

```

XX
XX      (IMMU-) IMMUNOLOGIC PHARM CORP.
PA
XX
XX      Gelfer ML, Garman RD, Greenstein JL, Bond JF;
PI
XX
XX      WPI; 2000-601477/57.
DR      N-PSDB; AAC60151.
XX
XX      Detecting, preventing and treating sensitivity to cat protein allergen
PT      comprises combining a biological sample with a human T cell reactive
PR      feline protein and determining the extent of binding that occurs -
PS      Disclosure; Figure 29; 106pp; English.
XX
XX      The present invention relates to the detection of sensitivity to a cat
CC      protein allergen by combining a blood sample from a subject with a
CC      peptide of human T cell reactive feline protein (hTRFP). This method
CC      and the hTRFP peptides are useful for diagnosing, preventing and
CC      treating cat allergies by reducing or abolishing an individual's
CC      allergic response to a cat allergen. DNA encoding the TRFP may be
CC      used as probes to locate equivalent sequences present in other species.
CC      These may further be used to study the mechanism of immunotherapy of
CC      cat allergy, and to design modified derivatives, analogues or
CC      functional equivalents useful in immunotherapy. The present
CC      sequence was used in the invention.
XX
SQ      Sequence 96 AA:

Query Match      23.4%; Score 26; DB 21; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33 FFAVANGNELLDLSITKYNATEPER 58
      ||||||||||||||||||||||||||||
DB      44 ffavangnellldlsitkynateper 69

RESULT 46
ID      AAY90145 standard; Protein; 96 AA.
XX
XX      AAY90145;
AC
XX
XX      13-JUL-2000 (first entry)
DT
XX
XX      TRFP fusion protein sequence.
DE
XX
XX      Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW      house dust; Fel d I; cat allergy; Fells domesticus sensitivity; therapy;
KW      diagnosis; goat; sheep; horse; rabbit; dog.
XX
XX      Synthetic.
OS
XX      US6025162-A.
PN
XX      15-FEB-2000.
PD
XX
XX      28-APR-1995; 95US-0430944.
PF
XX
XX      02-SEP-1994; 94US-0300928.
PR      03-NOV-1989; 89US-0431565.
PR      28-FEB-1991; 91US-0662276.
PR      13-DEC-1991; 91US-0807529.
PR      25-MAR-1992; 92US-0857311.
PR      15-MAY-1992; 92US-0884718.
PR      15-JAN-1993; 93US-0006116.
XX
XX      (IMMU-) IMMUNOLOGIC PHARM CORP.
PA
XX      Morgenstern JP, Griffith ID, Rogers BL;
PI
XX      WPI; 2000-181812/16.
DR

```

DR N-PSDB: AAA07480.  
XX  
XX New human T cell reactive feline protein, useful for desensitizing cat  
PT allergic individuals to cat allergens -  
XX  
XX  
XX Disclosure: Fig 29; 108pp; English.  
XX  
XX This sequence is a chimeric TRFP protein sequence. The invention  
CC relates to a human T cell reactive feline protein (TRFP). The protein is  
CC a cat protein allergen, and was isolated from a vacuum bag extract  
CC obtained by affinity purification of house dust collected from several  
CC homes with cats. TRFP is composed of two covalently linked peptide  
CC chains, and is also referred to as Fel d 1 TRFP and its peptides are  
CC useful for reducing or preventing the adverse effects that exposure  
CC to cat allergens normally has on cat allergic individuals (i.e.,  
CC desensitize individuals to cat allergens or block the effect of the  
CC allergens). TRFP is also used in methods of diagnosing sensitivity to  
CC fells domesticus in an individual. DNA sequences encoding TRFP can be  
CC used as probes to locate equivalent sequences present in other species,  
CC e.g. goat, sheep, horse, rabbit and dog, that may be useful in diagnostic  
CC and/or therapeutic applications.  
XX  
XX Sequence 96 AA:  
SQ  
Query Match 23.4%; Score 26; DB 21; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1,4e-17;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 33 FFVAVANGNELLDLSITKVNATEPER 58  
DB 44 ffavangnellldlsitkvateper 69  
RESULT 47  
AAB28962  
ID AAB28962 standard; Peptide; 27 AA.  
XX  
XX AAB28962;  
AC  
XX  
XX 29-JAN-2001 (first entry)  
DT  
XX  
XX TRFP chain 2 sequence Fel 18.  
DE  
XX  
XX Cat; allergy; human T cell reactive feline protein; hTRFP;  
KW immunotherapy.  
KM  
XX  
XX Fells sp.  
OS  
XX  
XX US6120769-A.  
PN  
XX  
XX 19-SEP-2000.  
PD  
XX  
XX 28-APR-1995; 95US-0431184.  
PF  
XX  
XX 02-SEP-1994; 94US-0300928.  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
PA  
XX  
XX Geffer ML, Garman RD, Greenstein JL, Bond JF;  
PI WPI; 2000-601477/57.  
XX  
XX  
XX Detecting, preventing and treating sensitivity to cat protein allergen  
PT comprises combining a biological sample with a human T cell reactive  
PT feline protein and determining the extent of binding that occurs -  
XX

PS Example 3; Column 34; 106pp; English.  
XX  
XX The present invention relates to the detection of sensitivity to a cat  
CC protein allergen by combining a blood sample from a subject with a  
CC peptide of human T cell reactive feline protein (hTRFP). This method  
CC and the hTRFP peptides are useful for diagnosing, preventing and  
CC treating cat allergies by reducing or abolishing an individual's  
CC allergic response to a cat allergen. DNA encoding the TRFP may be  
CC used as probes to locate equivalent sequences present in other species.  
CC These may further be used to study the mechanism of immunotherapy of  
CC cat allergy, and to design modified derivatives, analogues or  
CC functional equivalents useful in immunotherapy. The present  
XX sequence was used in the invention.  
XX  
XX Sequence 27 AA:  
SQ  
Query Match 20.7%; Score 23; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 42 LLDLSITKVNATEPERTAMKTI 64  
DB 1 lldlsitkvatepertamk1 23  
RESULT 48  
AAB7705  
ID AAB7705 standard; Protein; 27 AA.  
XX  
XX AAB7705;  
AC  
XX  
XX 22-AUG-2000 (first entry)  
DT  
XX  
XX Feline human TRFP peptide Fel 18.  
DE  
XX  
XX T-cell reactive feline protein; TRFP; Fel d 1; cat allergen;  
KW anti-allergic; T cell stimulator; diagnostic; immunotherapy.  
KM  
XX  
XX Fells sp.  
OS  
XX  
XX US6048962-A.  
PN  
XX  
XX 11-APR-2000.  
PD  
XX  
XX 27-APR-1995; 95US-0430014.  
PF  
XX  
XX 02-SEP-1994; 94US-0300928.  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
XX (IMMU-) IMMUNOLOGIC PHARM CORP.  
PA  
XX  
XX Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;  
PI Greenstein JL, Griffith ID, Garman RD;  
PI WPI; 2000-316905/27.  
XX  
XX New human T cell reactive feline protein useful for reducing or  
PT abolishing individual's allergic response to cat allergen comprising  
PT two different covalently linked peptide chains -  
XX  
XX Example 4; Column 103-104; 106pp; English.  
XX  
XX This invention describes a novel naturally occurring cat protein allergen  
CC (I) human T cell reactive feline protein (TRFP), comprising two  
CC different covalently linked peptide chains with a molecular weight of 20  
CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
CC under reducing conditions. The products of the invention have

CC antiallergic activity and act as human T cell stimulators. TRPP is useful  
CC for reducing or preventing the adverse effects of cat allergens on cat  
CC allergic individuals and in ex vivo diagnostic tests to determine which  
CC peptides cause sensitivity so as to selectively use them to desensitize  
CC a cat sensitive individual. Purified TRPP is also useful for studying  
CC the mechanism of immunotherapy of cat allergy and to design modified  
CC derivatives, analogs or functional equivalents that are more useful in  
CC immunotherapy against cat allergy. DNA sequences encoding TRPP are  
CC useful as probes to locate equivalent sequences present in other species  
CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
CC and/or therapeutics. Fully defined and characterized TRPP provides  
CC complete and a very simple desensitization therapy. This sequence  
CC represents a human T cell reactive feline protein (also known as Fel d I)  
CC derived peptide Fel 18 which is described in the method of the invention.  
XX  
SQ Sequence 27 AA;  
  
Query Match 20.7%; Score 23; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 42 LLLDLSITKYNATEPERTANKKI 64  
DB 1 LLLDLSITKYNATEPERTANKKI 23  
  
RESULT 49  
AAV90117  
ID AAV90117 standard; Peptide; 27 AA.  
XX  
AC AAV90117;  
XX  
DT 13-JUL-2000 (first entry)  
XX  
DE Cat TRPP chain 2 derived peptide, peptide Fel 18.  
XX  
KM Cat; TRPP; human T-cell reactive feline protein; cat protein allergen;  
KM house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;  
KM diagnosis; goat; sheep; horse; rabbit; dog.  
XX  
OS Felis domesticus.  
XX  
PN US6025162-A.  
XX  
PD 15-FEB-2000.  
XX  
PF 28-APR-1995; 950S-0430944.  
XX  
PR 02-SEP-1994; 94US-0300928.  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Morgenstern JP, Griffith TJ, Rogers BL;  
XX  
DR WPI; 2000-181812/16.  
XX  
PT New human T cell reactive feline protein, useful for desensitizing cat  
PT allergic individuals to cat allergens -  
XX  
PS Example 4; Column 35; 108bp; English.  
XX  
CC This sequence is a peptide derived from the human T cell reactive feline  
CC protein (TRPP) of the invention. The protein is a cat protein allergen,  
CC and was isolated from a vacuum bag extract obtained by affinity  
CC purification of house dust collected from several homes with cats. TRPP  
CC is composed of two covalently linked peptide chains, and is also referred

CC to as Fel d I. TRPP and its peptides are useful for reducing or  
CC preventing the adverse effects that exposure to cat allergens normally  
CC has on cat allergic individuals (i.e. to desensitize individuals to cat  
CC allergens or block the effect of the allergens). TRPP is also used in  
CC methods of diagnosing sensitivity to Felis domesticus in an individual.  
CC DNA sequences encoding TRPP can be used as probes to locate equivalent  
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and  
CC dog, that may be useful in diagnostic and/or therapeutic applications.  
XX  
SQ Sequence 27 AA;  
  
Query Match 20.7%; Score 23; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 42 LLLDLSITKYNATEPERTANKKI 64  
DB 1 LLLDLSITKYNATEPERTANKKI 23  
  
RESULT 50  
AAV51502  
ID AAV51502 standard; Protein; 27 AA.  
XX  
AC AAV51502;  
XX  
DT 22-MAY-2000 (first entry)  
XX  
DE Human TRPP derived peptide Fel 18.  
XX  
KM T-cell reactive feline protein; TRPP; T cell epitope; T cell receptor;  
KM down regulation; immune response; allergen; immunoglobulin E;  
KM sensitivity; cat protein allergen; human.  
XX  
OS Homo sapiens.  
XX  
PN US6019972-A.  
XX  
PD 01-FEB-2000.  
XX  
PF 02-SEP-1994; 94US-0300928.  
XX  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;  
XX  
DR WPI; 2000-146862/13.  
XX  
PT Peptides of human T cell reactive feline protein for treating  
PT sensitivity to cat protein allergens comprise at least one T cell  
PT epitope recognized by a T cell receptor specific for the human T cell  
PT reactive feline protein -  
XX  
PS Example 4; Column 103-104; 105bp; English.  
XX  
CC This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94,  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRPP

CC derived peptide used in the method of the invention.

XX Sequence 27 AA;

Query Match 20.7%, Score 23; DB 21; Length 27;  
Best Local Similarity 100.0%; Pred. No. 3.3e-15;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 42 LLDLSLTQVATEPERTAMKRI 64  
1 IIdsltkvnapertamkkl 23

RESULT 51

ID AAB28953 standard; Peptide: 20 AA.

AC AAB28953;

DT 29-JAN-2001 (first entry)

DE Peptide Fel 38.

KW Cat: allergy; human T cell reactive feline protein; hTRFP;  
immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 9505-0431184.

PR 02-SEP-1994; 9405-0300928.

PR 03-NOV-1989; 8905-0431565.

PR 28-FEB-1991; 9105-0662276.

PR 13-DEC-1991; 9105-0807529.

PR 25-MAR-1992; 9205-0857311.

PR 15-MAY-1992; 9205-0884718.

PR 15-JAN-1993; 9305-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Gelfer ML, Garman RD, Greenstein JL, Bond JF;

WP1: 2000-601477/57.

XX Detecting, preventing and treating sensitivity to cat protein allergen  
XX comprises combining a biological sample with a human T cell reactive  
XX feline protein and determining the extent of binding that occurs  
XX  
XX Example 7; Figure 18; 106pp; English.  
XX The present invention relates to the detection of sensitivity to a cat  
XX protein allergen by combining a blood sample from a subject with a  
XX peptide of human T cell reactive feline protein (hTRFP). This method  
XX and the hTRFP peptides are useful for diagnosing, preventing and  
XX treating cat allergies by reducing or abolishing an individual's  
XX allergic response to a cat allergen. DNA encoding the TRFP may be  
XX used as probes to locate equivalent sequences present in other species.  
XX These may further be used to study the mechanism of immunotherapy of  
XX cat allergy, and to design modified derivatives, analogues or  
XX functional equivalents useful in immunotherapy. The present  
XX sequence was used in the invention.

CC Sequence 20 AA;

Query Match 18.0%; Score 20; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ISRVLDGLVMTTSSKDCM 93  
1 ISrVldglvmttsskdcM 20

RESULT 52

ID AAY87694 standard; Protein: 20 AA.

AC AAY87694;

DT 22-AUG-2000 (first entry)

DE Feline human TRFP peptide Fel 38.

KW T-cell reactive feline protein; TRFP; Fel d 1; cat allergen;  
antiallergic; T cell stimulator; diagnostic; immunotherapy.

OS Felis sp.

PN US6048962-A.

PD 11-APR-2000.

PF 27-APR-1995; 9505-0430014.

PR 02-SEP-1994; 9405-0300928.

PR 03-NOV-1989; 8905-0431565.

PR 28-FEB-1991; 9105-0662276.

PR 13-DEC-1991; 9105-0807529.

PR 25-MAR-1992; 9205-0857311.

PR 15-MAY-1992; 9205-0884718.

PR 15-JAN-1993; 9305-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Kuo M, Rogers BL, Gelfer ML, Morgenstern JP, Brauer AM;

Greenstein JL, Griffith IT, Garman RD;

WP1: 2000-316905/27.

XX New human T cell reactive feline protein useful for reducing or  
XX abolishing individual's allergic response to cat allergen comprising  
XX two different covalently linked peptide chains  
XX  
XX Example 5; Column 91-92; 106pp; English.  
XX This invention describes a novel naturally occurring cat protein allergen  
XX (1), human T cell reactive feline protein (TRFP), comprising two  
XX different covalently linked peptide chains with a molecular weight of 20  
XX kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
XX under reducing conditions. The products of the invention have  
XX antiallergic activity and act as human T cell stimulators. TRFP is useful  
XX for reducing or preventing the adverse effects of cat allergens on cat  
XX allergic individuals and in ex vivo diagnostic tests to determine which  
XX peptides cause sensitivity so as to selectively use them to desensitize  
XX a cat sensitive individual. Purified TRFP is also useful for studying  
XX the mechanism of immunotherapy of cat allergy and to design modified  
XX derivatives, analogs or functional equivalents that are more useful in  
XX immunotherapy against cat allergy. DNA sequences encoding TRFP are  
XX useful as probes to locate equivalent sequences present in other species  
XX (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
XX and/or therapeutics. Fully defined and characterized TRFP provides  
XX complete and a very simple desensitization therapy. This sequence  
XX represents a human T cell reactive feline protein (also known as Fel d 1)  
XX derived peptide Fel 38 which is described in the method of the invention.

CC Sequence 20 AA;

Query Match 18.0%; Score 20; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ISRVLDGLVMTTSSSKDCM 93  
 ||||||||||||||||||  
 Db 1 Isrvldglvmttssskdcm 20

## RESULT 53

AA90113  
 ID AA90113 standard; Peptide: 20 AA.

XX  
 AC AA90113;

DT 13-JUL-2000 (first entry)

XX Cat TRFP derived peptide; peptide Fel 38.

XX Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;

KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;

KW diagnosis; goat; sheep; horse; rabbit; dog.

XX Felis domesticus.

XX US6025162-A.

PD 15-FEB-2000.

XX 28-APR-1995; 95US-0430944.

XX 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX PI Morgenstern JP, Griffith ID, Rogers BL;

XX WPI; 2000-181812/16.

XX New human T cell reactive feline protein, useful for desensitizing cat

PT allergic individuals to cat allergens -

XX Claim 9; Fig 18; 108pp; English.

XX This sequence is a peptide derived from the human T cell reactive feline

CC protein (TRFP) of the invention. The protein is a cat protein allergen,

CC and was isolated from a vacuum bag extract obtained by affinity

CC purification of house dust collected from several homes with cats. TRFP

CC is composed of two covalently linked peptide chains, and is also referred

CC to as Fel d I. TRFP and its peptides are useful for reducing or

CC preventing the adverse effects that exposure to cat allergens normally

CC has on cat allergic individuals (i.e. to desensitize individuals to cat

CC allergens or block the effect of the allergens). TRFP is also used in

CC methods of diagnosing sensitivity to Felis domesticus in an individual.

CC DNA sequences encoding TRFP can be used as probes to locate equivalent

CC sequences present in other species, e.g. goat, sheep, horse, rabbit and

CC dog, that may be useful in diagnostic and/or therapeutic applications.

XX Sequence 20 AA;

SO Query Match 18.0%; Score 20; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 2e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ISRVLDGLVMTTSSSKDCM 93

||||||||||||||||||

Db 1 Isrvldglvmttssskdcm 20

## RESULT 54

AA51491  
 ID AA51491 standard; Protein: 20 AA.

XX  
 AC AA51491;

DT 22-MAY-2000 (first entry)

XX Human TRFP derived peptide Fel 38.

XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;

KW down regulation; immune response; allergen; immunoglobulin E;

KW sensitivity; cat protein allergen; human.

XX Homo sapiens.

XX US6019972-A.

PD 01-FEB-2000.

XX 02-SEP-1994; 94US-0300928.

XX 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;

XX WPI; 2000-146862/13.

XX peptides of human T cell reactive feline protein for treating

PT sensitivity to cat protein allergens comprise at least one T cell

PT epitope recognized by a T cell receptor specific for the human T cell

PT reactive feline protein -

XX Example 5; Column 93-94; 105pp; English.

XX This invention describes a novel peptide (I) of human T cell reactive

CC feline protein (hTRFP) having at least one T cell epitope recognized

CC by a T cell receptor specific for the human T cell reactive feline

CC protein, the peptide consisting of at least 7-30 amino acids, and having

CC an amino acid sequence derived from an amino acid sequence comprising 94,

CC 96, 97, 109, or 111 residues, given in the specification. The peptides

CC down regulate the immune response to the allergen. The peptides have

CC reduced immunoglobulin E binding and reduce T cell responsiveness. The

CC peptide (I) is useful in compositions for treating sensitivity to a cat

CC protein allergen in a subject. This sequence represents the human TRFP

CC derived peptide used in the method of the invention.

XX Sequence 20 AA;

SO Query Match 18.0%; Score 20; DB 21; Length 20;

Best Local Similarity 100.0%; Pred. No. 2e-12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ISRVLDGLVMTTSSSKDCM 93

||||||||||||||||||

Db 1 Isrvldglvmttssskdcm 20

RESULT 55

AA36546

ID AA36546 standard; Protein: 19 AA.

XX

AC AA36546;

XX

```

DR 12-AUG-1993 (first entry)
XX
DE Peptide B.
XX
KW Human T cell reactive feline protein; TRFP; epitope: recombinant
XX sensitivity; Felis domesticus.
XX
OS Felis.
XX
PN W09308280-A.
XX
PD 29-APR-1993.
XX
PF 16-OCT-1992; 92MO-US08694.
XX
PR 16-OCT-1991; 91US-0777859.
XX
PR 13-DEC-1991; 91US-0807529.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
XX Rogers BL;
XX WPI; 1993-152473/18.
XX
DR
XX
PT Recombinant peptide having T-cell stimulating activity - for the
PT diagnosis and treatment of sensitivity to protein allergens,
PT autoantigens and protein antigens
XX
PS Disclosure: Fig 4; 73pp; English.
XX
CC Preferred recombinant peptides for treating sensitivity to Felis
CC domesticus are derived from the the genus Felis and comprise
CC regions selected from peptides X, Y, Z, A and B, of TRFP, and
CC modifications thereof, such as peptide C.
XX
SQ Sequence 19 AA:

Query Match 17.1%; Score 19; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 MGEAVONTVEDLKINTLGR 111
   |||
DB 1 mgeavqntvedlkintlgr 19

RESULT 56
AAR41979
ID AAR41979 standard; peptide; 19 AA.
XX
AC AAR41979;
XX
DT 21-APR-1994 (first entry)
XX
DE Human T cell reactive feline protein fragment B.
XX
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemisia;
KW Plantago; Parietaria; Blatella; Apis; Periplaneta; autoantigen; ss.
XX
OS Homo sapiens.
XX
PN W09319178-A.
XX
PD 30-SEP-1993.
XX
PF 25-MAR-1993; 93MO-US02462.
XX
PR 25-MAR-1992; 92US-0857311.
XX
PR 15-MAY-1992; 92US-0884718.

```

```

PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Briner TJ, Garman RD, Gefter ML, Greenstein JL;
XX Kuo M, Morville M;
XX WPI; 1993-320744/40.
XX
DR
XX
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
XX
PS Claim 1; Fig 3; 107pp; English.
XX
CC The sequences given in AAR41975-82 are peptides derived from a human T
CC cell reactive feline protein. These peptides are used in a
CC therapeutic composition which is useful in treating diseases which
CC involve an immune response to a protein antigen. This composition
CC may be used to induce tolerance in a mammal to Dermatophagoides,
CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
CC Quercus, Olea, Artemisia, Plantago, Parietaria, Canis, Blatella,
CC Apis, Periplaneta and to autoantigens in humans.
XX
SQ Sequence 19 AA:

Query Match 17.1%; Score 19; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 MGEAVONTVEDLKINTLGR 111
   |||
DB 1 mgeavqntvedlkintlgr 19

RESULT 57
AAB28945
ID AAB28945 standard; peptide; 19 AA.
XX
AC AAB28945;
XX
DT 29-JAN-2001 (first entry)
XX
DE Peptide B derived from T cell reactive feline protein.
XX
KW Cat; allergy; human T cell reactive feline protein; hTRFP;
KW immunotherapy.
XX
OS Felis sp.
XX
PN US6120769-A.
XX
PD 19-SEP-2000.
XX
PF 28-APR-1995; 95US-0431184.
XX
PR 02-SEP-1994; 94US-0300928.
XX
PR 03-NOV-1989; 89US-0431565.
XX
PR 28-FEB-1991; 91US-0662276.
XX
PR 13-DEC-1991; 91US-0807529.
XX
PR 25-MAR-1992; 92US-0857311.
XX
PR 15-MAY-1992; 92US-0884718.
XX
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Gefter ML, Garman RD, Greenstein JL, Bond JF;
XX WPI; 2000-601477/57.
XX
DR
XX
PT Detecting, preventing and treating sensitivity to cat protein allergen
PT comprises combining a biological sample with a human T cell reactive

```



xx feline protein and determining the extent of binding that occurs

ps Claim 2; Figure 17; 106bp; English.

cc The present invention relates to the detection of sensitivity to a cat  
cc protein allergen by combining a blood sample from a subject with a  
cc peptide of human T cell reactive feline protein (hTRFP). This method  
cc and the hTRFP peptides are useful for diagnosing, preventing and  
cc treating cat allergies by reducing or abolishing an individual's  
cc allergic response to a cat allergen. DNA encoding the TRFP may be  
cc used as probes to locate equivalent sequences present in other species.  
cc These may further be used to study the mechanism of immunotherapy of  
cc cat allergy, and to design modified derivatives, analogues or  
cc functional equivalents useful in immunotherapy. The present  
cc sequence was used in the invention.

xx  
xx Sequence 19 AA:

xx  
xx

xx Query Match 17.1%; Score 19; DB 21; Length 19;  
xx Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
xx Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

xx 93 MGEAVNTVEEDKIKTLGR 111  
xx ||||||||||||||||  
xx 1 mgeavntveedkikntlgr 19

xx

xx RESULT 58  
xx AAY87686  
xx ID AAY87686 standard; Protein: 19 AA.  
xx  
xx AAY87686;  
xx  
xx 22-AUG-2000 (first entry)  
xx  
xx Feline human TRFP peptide B.  
xx  
xx T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
xx antiallergic; T cell stimulator; diagnostic; immunotherapy.  
xx  
xx Felis sp.  
xx  
xx US6048962-A.  
xx  
xx 11-APR-2000.  
xx  
xx 27-APR-1995; 95US-0430014.  
xx  
xx 02-SEP-1994; 94US-0300928.  
xx 03-NOV-1989; 89US-0431565.  
xx 28-FEB-1991; 91US-0662276.  
xx 13-DEC-1991; 91US-0807529.  
xx 25-MAR-1992; 92US-0857311.  
xx 15-MAR-1992; 92US-0884718.  
xx 15-JAN-1993; 93US-0006116.  
xx  
xx (IMMU-) IMMUNOLOGIC PHARM CORP.  
xx  
xx Kuo M. Rogers BU, Gelfer ML, Morgenstern JP, Brauer AM;  
xx Greenstein JL, Griffith IV, Garman RD;  
xx WPI; 2000-316905/27.  
xx  
xx New human T cell reactive feline protein useful for reducing or  
xx abolishing individual's allergic response to cat allergen comprising  
xx two different covalently linked peptide chains  
xx  
xx Example 5; Column 87-88; 106bp; English.

cc This invention describes a novel naturally occurring cat protein allergen  
cc (1), human T cell reactive feline protein (TRFP), comprising two  
cc different covalently linked peptide chains with a molecular weight of 20

CC	kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
CC	under reducing conditions. The products of the invention have
CC	antiallergic activity and act as human T cell stimulators. TRFP is useful
CC	for reducing or preventing the adverse effects of cat allergens on cat
CC	allergic individuals and in ex vivo diagnostic tests to determine which
CC	peptides cause sensitivity so as to selectively use them to desensitize
CC	a cat sensitive individual. Purified TRFP is also useful for studying
CC	the mechanism of immunotherapy of cat allergy and to design modified
CC	derivatives, analogs or functional equivalents that are more useful in
CC	immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC	useful as probes to locate equivalent sequences present in other species
CC	(goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC	and/or therapeutics. Fully defined and characterized TRFP provides
CC	complete and a very simple desensitization therapy. This sequence
CC	represents a human T cell reactive feline protein (also known as Fel d I)
CC	derived peptide B which is described in the method of the invention.
XX	
SQ	Sequence 19 AA:
Query Match	17.1%; Score 19; DB 21; Length 19;
Best Local Similarity	100.0%; Pred. No. 1.8e-11;
Matches 19; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	93 MGEAVQNTVEDLKLTIGR 111
Db	I mgeavqntvedlkinltigr 19
RESULT 59	
AAV90111	
ID AAV90111 standard; Peptide: 19 AA.	
XX AAV90111;	
DT 13-JUL-2000 (first entry)	
DE Cat TRFP derived peptide, peptide B.	
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;	
KM house dust; Fel d I; cat allergy; feline domesticus sensitivity; therapy;	
KW diagnosis; goat; sheep; horse; rabbit; dog.	
XX Felis domesticus.	
OS US6025162-A.	
PN 15-FEB-2000.	
XX 28-APR-1995; 95US-0430944.	
XX 02-SEP-1994; 94US-0300928.	
PR 03-NOV-1989; 89US-0431565.	
PR 28-FEB-1991; 91US-0662276.	
PR 13-DEC-1991; 91US-0807529.	
PR 25-MAR-1992; 92US-0857311.	
PR 15-MAY-1992; 92US-0884718.	
PR 15-JAN-1993; 93US-0006116.	
PA (IMMU-) IMMUNOLOGIC PHARM CORP.	
PI Morgenstern JP, Griffith IU, Rogers BU,	
DR WPI; 2000-181812/16.	
XX New human T cell reactive feline protein, useful for desensitizing cat	
XX allergic individuals to cat allergens -	
PS Claim 9; Fig 17; 108pp; English.	
CC This sequence is a peptide derived from the human T cell reactive feline	
CC protein (TRFP) of the invention. The protein is a cat protein allergen,	
CC and was isolated from a vacuum bag extract obtained by affinity	

CC purification of house dust collected from several homes with cats. TRFP  
CC is composed of two covalently linked peptide chains, and is also referred  
CC to as Fel d I. TRFP and its peptides are useful for reducing or  
CC preventing the adverse effects that exposure to cat allergens normally  
CC has on cat allergic individuals (i.e. to desensitise individuals to cat  
CC allergens or block the effect of the allergens). TRFP is also used in  
CC methods of diagnosing sensitivity to feline allergens in an individual.  
CC DNA sequences encoding TRFP can be used as probes to locate equivalent  
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and  
CC dog, that may be useful in diagnostic and/or therapeutic applications.

SO Sequence 19 AA:

Query Match 17.1%; Score 19; DB 21; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 93 MGEAVQNTVEDKLNLTGR 111  
Db 1 mgeavqntvedklnltgr 19  
|||||

RESULT 60

AAV51483  
ID AAV51483 standard; Protein; 19 AA.

AC AAV51483;

DT 22-MAY-2000 (first entry)

DE Human TRFP derived peptide B.

KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;

KM down regulation; immune response; allergen; immunoglobulin E;

KX sensitivity; cat protein allergen; human.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;

DR WPI: 2000-146862/13.

XX Claim 8; Column 89-90; 105pp; English.

CC This invention describes a novel peptide (I) of human T cell reactive

CC feline protein (hTRFP) having at least one T cell epitope recognized

CC by a T cell receptor specific for the human T cell reactive feline

CC protein, the peptide consisting of at least 7-30 amino acids, and having

CC an amino acid sequence derived from an amino acid sequence comprising 94,

CC 96, 97, 109, or 111 residues, given in the specification. The peptides

CC down regulate the immune response to the allergen. The peptides have

CC reduced immunoglobulin E binding and reduce T cell responsiveness. The

CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC derived peptide used in the method of the invention.

SO Sequence 19 AA:

Query Match 17.1%; Score 19; DB 21; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.8e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 93 MGEAVQNTVEDKLNLTGR 111  
Db 1 mgeavqntvedklnltgr 19  
|||||

RESULT 61

AAB28949  
ID AAB28949 standard; Peptide; 18 AA.

AC AAB28949;

DT 29-JAN-2001 (first entry)

DE Peptide Fel 34.

KW Cat; allergy; human T cell reactive feline protein; hTRFP;

KM immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 95US-0431184.

PR 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

PA (IMMU-) IMMUNOLOGIC PHARM CORP.

PI Gelfer ML, Garman RD, Greenstein JL, Bond JF;

DR WPI: 2000-601477/57.

XX Example 7; Figure 18; 106pp; English.

CC The present invention relates to the detection of sensitivity to a cat

CC protein allergen by combining a blood sample from a subject with a

CC peptide of human T cell reactive feline protein (hTRFP). This method

CC and the hTRFP peptides are useful for diagnosing, preventing and

CC treating cat allergies by reducing or abolishing an individual's

CC allergic response to a cat allergen. DNA encoding the TRFP may be

CC used as probes to locate equivalent sequences present in other species.

CC These may further be used to study the mechanism of immunotherapy of

CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present

CC sequence was used in the invention.

SO Sequence 18 AA:

Query Match 16.2%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 NGNELLDLSLTKVNATE 55  
|  
Db 1 ngnellldlsltkvnate 18

## RESULT 62

AAB28950  
ID AAB28950 standard; Peptide; 18 AA.

AC AAB28950;

DT 29-JAN-2001 (first entry)

DE Peptide Fel 35.

KM Cat; allergy; human T cell reactive feline protein; hTRFP;

KW Immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 950S-0431184.

PR 02-SEP-1994; 940S-0300928.

PR 03-NOV-1989; 890S-0431565.

PR 28-FEB-1991; 910S-0662276.

PR 13-DEC-1991; 910S-0807529.

PR 25-MAR-1992; 920S-0857311.

PR 15-MAY-1992; 920S-0884718.

PR 15-JAN-1993; 930S-0006116.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Gefter ML, Garman RD, Greenstein JL, Bond JF;

DR WPI; 2000-601477/57.

XX

PT Detecting, preventing and treating sensitivity to cat protein allergen

PT comprises combining a biological sample with a human T cell reactive

PT feline protein and determining the extent of binding that occurs.

XX

PS Example 7; Figure 18; 106pp; English.

XX

CC The present invention relates to the detection of sensitivity to a cat

CC protein allergen by combining a blood sample from a subject with a

CC peptide of human T cell reactive feline protein (hTRFP). This method

CC and the hTRFP peptides are useful for diagnosing, preventing and

CC treating cat allergies by reducing or abolishing an individual's

CC allergic response to a cat allergen. DNA encoding the TRFP may be

CC used as probes to locate equivalent sequences present in other species.

CC These may further be used to study the mechanism of immunotherapy of

CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present

CC sequence was used in the invention.

CC

XX

SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.6e-10;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 SLTKVNATEPERTAMKKI 64  
|  
Db 1 sltkvnatepertamkki 18

## RESULT 63

AAB28952  
ID AAB28952 standard; Peptide; 18 AA.

AC AAB28952;

DT 29-JAN-2001 (first entry)

DE Peptide Fel 37.

KM Cat; allergy; human T cell reactive feline protein; hTRFP;

KW Immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 950S-0431184.

PR 02-SEP-1994; 940S-0300928.

PR 03-NOV-1989; 890S-0431565.

PR 28-FEB-1991; 910S-0662276.

PR 13-DEC-1991; 910S-0807529.

PR 25-MAR-1992; 920S-0857311.

PR 15-MAY-1992; 920S-0884718.

PR 15-JAN-1993; 930S-0006116.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Gefter ML, Garman RD, Greenstein JL, Bond JF;

DR WPI; 2000-601477/57.

XX

PT Detecting, preventing and treating sensitivity to cat protein allergen

PT comprises combining a biological sample with a human T cell reactive

PT feline protein and determining the extent of binding that occurs.

XX

PS Example 7; Figure 18; 106pp; English.

XX

CC The present invention relates to the detection of sensitivity to a cat

CC protein allergen by combining a blood sample from a subject with a

CC peptide of human T cell reactive feline protein (hTRFP). This method

CC and the hTRFP peptides are useful for diagnosing, preventing and

CC treating cat allergies by reducing or abolishing an individual's

CC allergic response to a cat allergen. DNA encoding the TRFP may be

CC used as probes to locate equivalent sequences present in other species.

CC These may further be used to study the mechanism of immunotherapy of

CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present

CC sequence was used in the invention.

CC

XX

SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.6e-10;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 QDCYVENGILSRVLDGLV 82  
|  
Db 1 qdcyvengilsrvldglv 18

## RESULT 64

AAB28982  
ID AAB28982 standard; Peptide; 18 AA.

AC AAB28982;

DT 29-JAN-2001 (first entry)

DE Peptide Fel 37.

KM Cat; allergy; human T cell reactive feline protein; hTRFP;

KW Immunotherapy.

OS Felis sp.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 950S-0431184.

PR 02-SEP-1994; 940S-0300928.

PR 03-NOV-1989; 890S-0431565.

PR 28-FEB-1991; 910S-0662276.

PR 13-DEC-1991; 910S-0807529.

PR 25-MAR-1992; 920S-0857311.

PR 15-MAY-1992; 920S-0884718.

PR 15-JAN-1993; 930S-0006116.

PA (IMMU-) IMMULOGIC PHARM CORP.

PI Gefter ML, Garman RD, Greenstein JL, Bond JF;

DR WPI; 2000-601477/57.

XX

PT Detecting, preventing and treating sensitivity to cat protein allergen

PT comprises combining a biological sample with a human T cell reactive

PT feline protein and determining the extent of binding that occurs.

XX

PS Example 7; Figure 18; 106pp; English.

XX

CC The present invention relates to the detection of sensitivity to a cat

CC protein allergen by combining a blood sample from a subject with a

CC peptide of human T cell reactive feline protein (hTRFP). This method

CC and the hTRFP peptides are useful for diagnosing, preventing and

CC treating cat allergies by reducing or abolishing an individual's

CC allergic response to a cat allergen. DNA encoding the TRFP may be

CC used as probes to locate equivalent sequences present in other species.

CC These may further be used to study the mechanism of immunotherapy of

CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present

CC sequence was used in the invention.

CC

XX

SQ Sequence 18 AA;

De Peptide Fel-32.

KW Cat; allergy; human T cell reactive feline protein; hTRFP;  
immunotherapy.

OS Fells sp.  
OS Synthetic.

KA US6120769-A.

PD 19-SEP-2000.

XX 28-APR-1995; 95US-0431184.

XX 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Geffer ML, Garman RD, Greenstein JL, Bond JF;

PI WPI; 2000-601477/57.

DR WPI; 2000-601477/57.

XX Detecting, preventing and treating sensitivity to cat protein allergen

PT comprises combining a biological sample with a human T cell reactive

PT feline protein and determining the extent of binding that occurs

XX Example 7: Column 105-106: 106pp: English.

XX The present invention relates to the detection of sensitivity to a cat

CC protein allergen by combining a blood sample from a subject with a

CC peptide of human T cell reactive feline protein (hTRFP). This method

CC and the hTRFP peptides are useful for diagnosing, preventing and

CC treating cat allergies by reducing or abolishing an individual's

CC allergic response to a cat allergen. DNA encoding the TRFP may be

CC used as probes to locate equivalent sequences present in other species.

CC These may further be used to study the mechanism of immunotherapy of

CC cat allergy, and to design modified derivatives, analogues or

CC functional equivalents useful in immunotherapy. The present

CC sequence was used in the invention.

XX Sequence 18 AA:

SQ

Query Match 16.2%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.6e-10;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 VKMAETCPFYDFEFAVA 37

DB 1 VKMAETCPFYDFEFAVA 18

RESULT 65

AAV87690

ID AAV87690 standard; Protein; 18 AA.

AC AAV87690;

DT 22-AUG-2000 (first entry)

XX Feline human TRFP peptide Fel 34.

XX T-cell reactive feline protein; TRFP; Fel d I; cat allergen;

XX anti-allergic; T cell stimulator; diagnostic; immunotherapy.

OS Fells sp.

PN US6048962-A.

PD 11-APR-2000.

XX 27-APR-1995; 95US-0430014.

XX 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 28-FEB-1991; 91US-0662276.

PR 13-DEC-1991; 91US-0807529.

PR 25-MAR-1992; 92US-0857311.

PR 15-MAY-1992; 92US-0884718.

PR 15-JAN-1993; 93US-0006116.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;

PI Greenstein JL, Griffith LJ, Garman RD;

DR WPI; 2000-316905/27.

XX New human T cell reactive feline protein useful for reducing or

PT abolishing individual's allergic response to cat allergen comprising

PT two different covalently linked peptide chains

XX Example 5: Column 89-90: 106pp: English.

XX This invention describes a novel naturally occurring cat protein allergen

CC (I), human T cell reactive feline protein (TRFP), comprising two

CC different covalently linked peptide chains with a molecular weight of 20

CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD

CC under reducing conditions. The products of the invention have

CC anti-allergic activity and act as human T cell stimulators. TRFP is useful

CC for reducing or preventing the adverse effects of cat allergens on cat

CC allergic individuals and in ex vivo diagnostic tests to determine which

CC peptides cause sensitivity so as to selectively use them to desensitize

CC a cat sensitive individual. Purified TRFP is also useful for studying

CC the mechanism of immunotherapy of cat allergy and to design modified

CC derivatives, analogs or functional equivalents that are more useful in

CC immunotherapy against cat allergy. DNA sequences encoding TRFP are

CC useful as probes to locate equivalent sequences present in other species

CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics

CC and/or therapeutics. Fully defined and characterized TRFP provides

CC complete and a very simple desensitization therapy. This sequence

CC represents a human T cell reactive feline protein (also known as Fel d I)

Sequence 18 AA:

Query Match 16.2%; Score 18; DB 21; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.6e-10;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 NGNELLDLSLTKVNATE 55

DB 1 ngnellldsltkvnate 18

RESULT 66

AAV87691

ID AAV87691 standard; Protein; 18 AA.

AC AAV87691;

DT 22-AUG-2000 (first entry)

XX Feline human TRFP peptide Fel 35.

XX T-cell reactive feline protein; TRFP; Fel d I; cat allergen;

XX anti-allergic; T cell stimulator; diagnostic; immunotherapy.

OS Fells sp.

```

XX  US6048962-A.
XX
XX  11-APR-2000.
XX
XX  27-APR-1995; 95US-0430014.
XX
XX  02-SEP-1994; 94US-0300928.
XX  03-NOV-1989; 89US-0431565.
XX  28-FEB-1991; 91US-0662276.
XX  13-DEC-1991; 91US-0807529.
XX  25-MAR-1992; 92US-0857311.
XX  15-MAY-1992; 92US-0884718.
XX  15-JAN-1993; 93US-0006116.
XX
XX  (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX  Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AM;
XX  Greenstein JL, Griffith IJ, Garman RD;
XX
XX  WPI; 2000-316905/27.
XX
XX  New human T cell reactive feline protein useful for reducing or
XX  abolishing individual's allergic response to cat allergen comprising
XX  two different covalently linked peptide chains
XX
XX  Example 5; Column 91-92; 106pp; English.
XX
XX  This invention describes a novel naturally occurring cat protein allergen
XX  (I), human T cell reactive feline protein (TRFP), comprising two
XX  different covalently linked peptide chains with a molecular weight of 20
XX  kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
XX  under reducing conditions. The products of the invention have
XX  antiallergic activity and act as human T cell stimulators. TRFP is useful
XX  for reducing or preventing the adverse effects of cat allergens on cat
XX  allergic individuals and in ex vivo diagnostic tests to determine which
XX  peptides cause sensitivity so as to selectively use them to desensitize
XX  a cat sensitive individual. Purified TRFP is also useful for studying
XX  the mechanism of immunotherapy of cat allergy and to design modified
XX  derivatives, analogs or functional equivalents that are more useful in
XX  immunotherapy against cat allergy. DNA sequences encoding TRFP are
XX  useful as probes to locate equivalent sequences present in other species
XX  (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
XX  and/or therapeutics. Fully defined and characterized TRFP provides
XX  complete and a very simple desensitization therapy. This sequence
XX  represents a human T cell reactive feline protein (also known as Fel d I)
XX  derived peptide Fel 35 which is described in the method of the invention.
XX
XX  Sequence 18 AA;
XX
SQ

```

```

Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 SLTKVNATEPERTAMKKI 64
Db 1 SLTKVNATEPERTAMKKI 18

```

```

RESULT 67
AA87693
ID AAY87693 standard; Protein; 18 AA.
XX
XX  AAY87693;
XX
XX  22-AUG-2000 (first entry)
XX
XX  Feline human TRFP peptide Fel 37.
XX
XX  T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
XX  antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX

```

```

OS  Fells sp.
XX
XX  US6048962-A.
XX
XX  11-APR-2000.
XX
XX  27-APR-1995; 95US-0430014.
XX
XX  02-SEP-1994; 94US-0300928.
XX  03-NOV-1989; 89US-0431565.
XX  28-FEB-1991; 91US-0662276.
XX  13-DEC-1991; 91US-0807529.
XX  25-MAR-1992; 92US-0857311.
XX  15-MAY-1992; 92US-0884718.
XX  15-JAN-1993; 93US-0006116.
XX
XX  (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX  Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AM;
XX  Greenstein JL, Griffith IJ, Garman RD;
XX
XX  WPI; 2000-316905/27.
XX
XX  New human T cell reactive feline protein useful for reducing or
XX  abolishing individual's allergic response to cat allergen comprising
XX  two different covalently linked peptide chains
XX
XX  Example 5; Column 91-92; 106pp; English.
XX
XX  This invention describes a novel naturally occurring cat protein allergen
XX  (I), human T cell reactive feline protein (TRFP), comprising two
XX  different covalently linked peptide chains with a molecular weight of 20
XX  kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
XX  under reducing conditions. The products of the invention have
XX  antiallergic activity and act as human T cell stimulators. TRFP is useful
XX  for reducing or preventing the adverse effects of cat allergens on cat
XX  allergic individuals and in ex vivo diagnostic tests to determine which
XX  peptides cause sensitivity so as to selectively use them to desensitize
XX  a cat sensitive individual. Purified TRFP is also useful for studying
XX  the mechanism of immunotherapy of cat allergy and to design modified
XX  derivatives, analogs or functional equivalents that are more useful in
XX  immunotherapy against cat allergy. DNA sequences encoding TRFP are
XX  useful as probes to locate equivalent sequences present in other species
XX  (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
XX  and/or therapeutics. Fully defined and characterized TRFP provides
XX  complete and a very simple desensitization therapy. This sequence
XX  represents a human T cell reactive feline protein (also known as Fel d I)
XX  derived peptide Fel 37 which is described in the method of the invention.
XX
XX  Sequence 18 AA;
XX
SQ

```

```

Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 65 ODCYVENGLISRYLDGLV 82
Db 1 qdcyvenglisryldglv 18

```

```

RESULT 68
AA87706
ID AAY87706 standard; Protein; 18 AA.
XX
XX  AAY87706;
XX
XX  22-AUG-2000 (first entry)
XX
XX  Feline human TRFP peptide Fel 32.
XX
XX  T-cell reactive feline protein; TRFP; Fel d I; cat allergen;
XX  antiallergic; T cell stimulator; diagnostic; immunotherapy.
XX

```

```

XX OS Fells sp.
XX KX US6048962-A.
XX PN
XX PD 11-APR-2000.
XX PF 27-APR-1995; 95US-0430014.
XX PR 02-SEP-1994; 94US-0300928.
XX PR 03-NOV-1989; 89US-0431565.
XX PR 28-FEB-1991; 91US-0662276.
XX PR 13-DEC-1991; 91US-0807529.
XX PR 25-MAR-1991; 92US-0857311.
XX PR 15-MAY-1992; 92US-0884718.
XX PR 15-JAN-1993; 93US-0006116.
XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX PI Kuo M, Rogers BL, Gelfert ML, Morgenstern JP, Brauer AM,
XX PI Greenstein JL, Griffith IU, Garman RD;
XX DR WPI: 2000-316905/27.
XX PS
XX PT New human T cell reactive feline protein useful for reducing or
XX PT abolishing individual's allergic response to cat allergen comprising
XX PT two different covalently linked peptide chains
XX PS
XX PS Example 7; Column 105-106; 106pp; English.
XX CC This invention describes a novel naturally occurring cat protein allergen
XX CC (I), human T cell reactive feline protein (TRFP), comprising two
XX CC different covalently linked peptide chains with a molecular weight of 20
XX CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD
XX CC under reducing conditions. The products of the invention have
XX CC antiallergic activity and act as human T cell stimulators. TRFP is useful
XX CC for reducing or preventing the adverse effects of cat allergens on cat
XX CC allergic individuals and in ex vivo diagnostic tests to determine which
XX CC peptides cause sensitivity so as to selectively desensitize
XX CC a cat sensitive individual. Purified TRFP is also useful for studying
XX CC the mechanism of immunotherapy of cat allergy and to design modified
XX CC derivatives, analogs or functional equivalents that are more useful in
XX CC immunotherapy against cat allergy. DNA sequences encoding TRFP are
XX CC useful as probes to locate equivalent sequences present in other species
XX CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
XX CC and/or therapeutics. Fully defined and characterized TRFP provides
XX CC complete and a very simple desensitization therapy. This sequence
XX CC represents a human T cell reactive feline protein (also known as Fel d 1)
XX CC derived peptide Fel 32 which is described in the method of the invention.
XX SO Sequence 18 AA:

```

Query Match 16.2%; Score 18; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 20 VKMAETCPIFYDVFFAVA 37
DB 1 VKmaetcpifydvffava 18

```

RESULT 69  
 ID AAY90118 standard; Peptide: 18 AA.  
 AC AAY90118;  
 XX  
 XX 13-JUL-2000 (first entry)  
 DE Cat TRFP derived peptide; peptide Fel 32.  
 XX  
 KM Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;

```

KW house dust; Fel d 1; cat allergy; Felis domesticus sensitivity; therapy;
XX diagnosis; goat; sheep; horse; rabbit; dog.
XX OS Fells domesticus.
XX KX US6025162-A.
XX PN
XX PD 15-FEB-2000.
XX PF 28-APR-1995; 95US-0430944.
XX PR 02-SEP-1994; 94US-0300928.
XX PR 03-NOV-1989; 89US-0431565.
XX PR 28-FEB-1991; 91US-0662276.
XX PR 13-DEC-1991; 91US-0807529.
XX PR 25-MAR-1991; 92US-0857311.
XX PR 15-MAY-1992; 92US-0884718.
XX PR 15-JAN-1993; 93US-0006116.
XX PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX PI Morgenstern JP, Griffith IU, Rogers BL;
XX PI Greenstein JL, Griffith IU, Garman RD;
XX DR WPI: 2000-181812/16.
XX PS
XX PT New human T cell reactive feline protein, useful for desensitizing cat
XX PT allergic individuals to cat allergens -
XX PS
XX PS Example 7; Fig 18; 108pp; English.
XX CC This sequence is a peptide derived from the human T cell reactive feline
XX CC protein (TRFP) of the invention. The protein is a cat protein allergen,
XX CC and was isolated from a vacuum bag extract obtained by affinity
XX CC purification of house dust collected from several homes with cats. TRFP
XX CC is composed of two covalently linked peptide chains, and is also referred
XX CC to as Fel d 1. TRFP and its peptides are useful for reducing or
XX CC preventing the adverse effects that exposure to cat allergens normally
XX CC has on cat allergic individuals (i.e. to desensitize individuals to cat
XX CC allergens or block the effect of the allergens). TRFP is also used in
XX CC methods of diagnosing sensitivity to fells domesticus in an individual.
XX CC DNA sequences encoding TRFP can be used as probes to locate equivalent
XX CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
XX CC dog, that may be useful in diagnostic and/or therapeutic applications.
XX SO Sequence 18 AA:

```

Query Match 16.2%; Score 18; DB 21; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 20 VKMAETCPIFYDVFFAVA 37
DB 1 VKmaetcpifydvffava 18

```

RESULT 70  
 ID AAY90120 standard; Peptide: 18 AA.  
 AC AAY90120;  
 XX  
 XX 13-JUL-2000 (first entry)  
 DE Cat TRFP derived peptide; peptide Fel 34.  
 XX  
 XX Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;  
 KM house dust; Fel d 1; cat allergy; Felis domesticus sensitivity; therapy;  
 XX diagnosis; goat; sheep; horse; rabbit; dog.  
 XX OS Fells domesticus.  
 XX KX US6025162-A.

```

XX 15-FEB-2000.
PD
XX
PF 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 23-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Morgenstern JP, Griffith IJ, Rogers BL;
XX
DR WPI; 2000-181812/16.
XX
PT New human T cell reactive feline protein, useful for desensitizing cat
PT allergic individuals to cat allergens -
XX
PS Example 7; Fig 18; 108pp; English.
XX
CC This sequence is a peptide derived from the human T cell reactive feline
CC protein (TRFP) of the invention. The protein is a cat protein allergen,
CC and was isolated from a vacuum bag extract obtained by affinity
CC purification of house dust collected from several homes with cats. TRFP
CC is composed of two covalently linked peptide chains, and is also referred
CC to as Fel d I. TRFP and its peptides are useful for reducing or
CC preventing the adverse effects that exposure to cat allergens normally
CC has on cat allergic individuals (i.e. to desensitize individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in diagnostic and/or therapeutic applications.
XX
SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 NGNELLDLSITKYNATE 55
Db 1 ngnelldlsitkynate 18

RESULT 71
AAV90121
ID AAV90121 standard; Peptide; 18 AA.
XX
AC AAV90121;
XX
DE 13-JUL-2000 (first entry)
XX
DE Cat TRFP derived peptide, peptide Fel 35.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PF 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.

```

```

PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 23-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
PI Morgenstern JP, Griffith IJ, Rogers BL;
XX
DR WPI; 2000-181812/16.
XX
DR N-PSDB; 3.
XX
PT New human T cell reactive feline protein, useful for desensitizing cat
PT allergic individuals to cat allergens -
XX
PS Example 7; Fig 18; 108pp; English.
XX
CC This sequence is a peptide derived from the human T cell reactive feline
CC protein (TRFP) of the invention. The protein is a cat protein allergen,
CC and was isolated from a vacuum bag extract obtained by affinity
CC purification of house dust collected from several homes with cats. TRFP
CC is composed of two covalently linked peptide chains, and is also referred
CC to as Fel d I. TRFP and its peptides are useful for reducing or
CC preventing the adverse effects that exposure to cat allergens normally
CC has on cat allergic individuals (i.e. to desensitize individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in diagnostic and/or therapeutic applications.
XX
SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 SLTKVNATEPERTAMKTI 64
Db 1 sltkvnatepertamk1 18

RESULT 72
AAV90123
ID AAV90123 standard; Peptide; 18 AA.
XX
AC AAV90123;
XX
DE 13-JUL-2000 (first entry)
XX
DE Cat TRFP derived peptide, peptide Fel 37.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PF 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 23-MAR-1992; 92US-0857311.
PR 15-MAY-1992; 92US-0884718.

```

```
PR 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Morgenstern JP, Griffith JJ, Rogers BL;
XX
XX WPI; 2000-181812/16.
XX
XX New human T cell reactive feline protein, useful for desensitizing cat
XX allergic individuals to cat allergens -
XX
XX Example 7; Fig 18; 108pp; English.
XX
XX This sequence is a peptide derived from the human T cell reactive feline
XX protein (TRFP) of the invention. The protein is a cat protein allergen,
XX and was isolated from a vacuum bag extract obtained by affinity
XX purification of house dust collected from several homes with cats. TRFP
XX is composed of two covalently linked peptide chains which are also referred
XX to as Fel d I. TRFP and its peptides are useful for reducing or
XX preventing the adverse effects that exposure to cat allergens normally
XX has on cat allergic individuals (i.e. to desensitize individuals to cat
XX allergens or block the effect of the allergens). TRFP is also used in
XX methods of diagnosing sensitivity to feline allergens in an individual.
XX DNA sequences encoding TRFP can be used as probes to locate equivalent
XX sequences present in other species, e.g. goat, sheep, horse, rabbit and
XX dog, that may be useful in diagnostic and/or therapeutic applications.
XX
XX Sequence 18 AA:
SQ
Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 65 QDCYVENGSLRWLDGLV 82
DB 1 qdcyvgngllsrivldgllv 18
RESULT 73
AAVS1487
AAVS1487 standard; Protein; 18 AA.
XX
XX AAVS1487;
XX
XX 22-MAY-2000 (first entry)
XX
XX Human TRFP derived peptide Fel 34.
XX
XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
XX down regulation; immune response; allergen; immunoglobulin E;
XX sensitivity; cat protein allergen; human.
XX
XX Homo sapiens.
XX
XX US6019972-A.
XX
XX 01-FEB-2000.
XX
XX 02-SEP-1994; 94US-0300928.
XX
XX 03-NOV-1989; 89US-0431565.
XX 28-FEB-1991; 91US-0662276.
XX 13-DEC-1991; 91US-0807529.
XX 25-MAR-1992; 92US-0857311.
XX 15-MAY-1992; 92US-0884718.
XX 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;
XX WPI; 2000-146862/13.
XX
XX
```

```
XX
XX Peptides of human T cell reactive feline protein for treating
XX sensitivity to cat protein allergens comprise at least one T cell
XX epitope recognized by a T cell receptor specific for the human T cell
XX reactive feline protein -
XX
XX Example 5; Column 91-92; 105pp; English.
XX
XX This invention describes a novel peptide (I) of human T cell reactive
XX feline protein (hTRFP) having at least one T cell epitope recognized
XX by a T cell receptor specific for the human T cell reactive feline
XX protein, the peptide consisting of at least 7-30 amino acids, and having
XX an amino acid sequence derived from an amino acid sequence comprising 94,
XX 96, 97, 109, or 111 residues, given in the specification. The peptides
XX down regulate the immune response to the allergen. The peptides have
XX reduced immunoglobulin E binding and reduce T cell responsiveness. The
XX peptide (I) is useful in compositions for treating sensitivity to a cat
XX protein allergen in a subject. This sequence represents the human TRFP
XX derived peptide used in the method of the invention.
XX
XX Sequence 18 AA:
SQ
Query Match 16.2%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 38 NGNELLDLSLRKNATE 55
DB 1 ngnellldlslrkvnate 18
RESULT 74
AAVS1488
AAVS1488 standard; Protein; 18 AA.
XX
XX AAVS1488;
XX
XX 22-MAY-2000 (first entry)
XX
XX Human TRFP derived peptide Fel 35.
XX
XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;
XX down regulation; immune response; allergen; immunoglobulin E;
XX sensitivity; cat protein allergen; human.
XX
XX Homo sapiens.
XX
XX US6019972-A.
XX
XX 01-FEB-2000.
XX
XX 02-SEP-1994; 94US-0300928.
XX
XX 03-NOV-1989; 89US-0431565.
XX 28-FEB-1991; 91US-0662276.
XX 13-DEC-1991; 91US-0807529.
XX 25-MAR-1992; 92US-0857311.
XX 15-MAY-1992; 92US-0884718.
XX 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;
XX WPI; 2000-146862/13.
XX
XX Peptides of human T cell reactive feline protein for treating
XX sensitivity to cat protein allergens comprise at least one T cell
XX epitope recognized by a T cell receptor specific for the human T cell
XX reactive feline protein -
XX
XX Example 5; Column 91-92; 105pp; English.
XX
XX
```



XX This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94,  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC derived peptide used in the method of the invention.

SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SLTKVNATEPERRTAMKKI 64  
|||||  
Db 1 sltkvnatepertamkkl 18

RESULT 75  
AAV51490  
ID AAV51490 standard; Protein; 18 AA.  
XX  
AC AAV51490;  
XX  
DT 22-MAY-2000 (first entry)  
XX  
DE Human TRFP derived peptide Fel 37.  
XX  
KM T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
KM down regulation; immune response; allergen; immunoglobulin E;  
KM sensitivity; cat protein allergen; human.  
XX  
OS Homo sapiens.  
XX  
PN US6019972-A.  
XX  
PD 01-FEB-2000.  
XX  
PF 02-SEP-1994; 94US-0300928.  
XX  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;  
XX  
DR WPI; 2000-146862/13.  
XX  
PT Peptides of human T cell reactive feline protein for treating  
PT sensitivity to cat protein allergens comprise at least one T cell  
PT epitope recognized by a T cell receptor specific for the human T cell  
PT reactive feline protein -  
XX  
PS Example 5; Column 91-92; 105pp; English.  
XX  
CC This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94,  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides

CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC derived peptide used in the method of the invention.

SQ Sequence 18 AA;

Query Match 16.2%; Score 18; DB 21; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 ODCYVENGISRYVDGLV 82  
|||||  
Db 1 qdcyvengisryvdglv 18

RESULT 76  
AAV51503  
ID AAV51503 standard; Protein; 18 AA.  
XX  
AC AAV51503;  
XX  
DT 22-MAY-2000 (first entry)  
XX  
DE Human TRFP derived peptide Fel 32.  
XX  
KM T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
KM down regulation; immune response; allergen; immunoglobulin E;  
KM sensitivity; cat protein allergen; human.  
XX  
OS Homo sapiens.  
XX  
PN US6019972-A.  
XX  
PD 01-FEB-2000.  
XX  
PF 02-SEP-1994; 94US-0300928.  
XX  
PR 03-NOV-1989; 89US-0431565.  
PR 28-FEB-1991; 91US-0662276.  
PR 13-DEC-1991; 91US-0807529.  
PR 25-MAR-1992; 92US-0857311.  
PR 15-MAY-1992; 92US-0884718.  
PR 15-JAN-1993; 93US-0006116.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gefter ML;  
XX  
DR WPI; 2000-146862/13.  
XX  
PT Peptides of human T cell reactive feline protein for treating  
PT sensitivity to cat protein allergens comprise at least one T cell  
PT epitope recognized by a T cell receptor specific for the human T cell  
PT reactive feline protein -  
XX  
PS Example 7; Column 105-106; 105pp; English.  
XX  
CC This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94,  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC derived peptide used in the method of the invention.

SQ Sequence 18 AA;



Db 1 mtltsskdcmgavqnt 18

RESULT 79

ID AAY90125 standard; Peptide: 30 AA.

XX AAY90125;

XX 13-JUL-2000 (first entry)

XX Cat TRFP derived peptide; peptide Fel 39.

XX Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;

XX house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;

XX diagnosis; goat; sheep; horse; rabbit; dog.

XX Felis domesticus.

XX US6025162-A.

XX 15-FEB-2000.

XX 28-APR-1995: 95US-0430944.

XX 02-SEP-1994: 94US-0300928.

XX 03-NOV-1989: 89US-0431565.

XX 28-FEB-1991: 91US-0662276.

XX 13-DEC-1991: 91US-0807529.

XX 25-MAR-1992: 92US-0857311.

XX 15-MAY-1992: 92US-0884718.

XX 15-JAN-1993: 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Morgenstern JP, Griffith JF, Rogers BL;

XX WPI: 2000-181812/16.

XX New human T cell reactive feline protein, useful for desensitizing cat

XX allergic individuals to cat allergens -

XX Example 7; Fig 18; 108pp; English.

XX This sequence is a peptide derived from the human T cell reactive feline

XX protein (TRFP) of the invention. The protein is a cat protein allergen,

XX and was isolated from a vacuum bag extract obtained by affinity

XX purification of house dust collected from several homes with cats. TRFP

XX is composed of two covalently linked peptide chains, and is also referred

XX to as Fel d I. TRFP and its peptides are useful for reducing or

XX preventing the adverse effects that exposure to cat allergens normally

XX has on cat allergic individuals (i.e. to desensitize individuals to cat

XX allergens or block the effect of the allergens). TRFP is also used in

XX methods of diagnosing sensitivity to feline domesticus in an individual.

XX DNA sequences encoding TRFP can be used as probes to locate equivalent

XX sequences present in other species, e.g. goat, sheep, horse, rabbit and

XX dog, that may be useful in diagnostic and/or therapeutic applications.

XX Sequence 30 AA;

Query Match 16.2%; Score 18; DB 21; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.7e-10; Mismatches 0; Indels 0; Gaps 0;

QY 83 MTTTSSSKDCMGAVONT 100

Db 1 mtltsskdcmgavqnt 18

RESULT 80

AAY51493

ID AAY51493 standard; Protein: 30 AA.

XX AAY51493;

XX 22-MAY-2000 (first entry)

XX Human TRFP derived peptide Fel 39.

XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;

XX down regulation; immune response; allergen; immunoglobulin E;

XX sensitivity; cat protein allergen; human.

XX Homo sapiens.

XX US6019972-A.

XX 01-FEB-2000.

XX 02-SEP-1994: 94US-0300928.

XX 03-NOV-1989: 89US-0431565.

XX 28-FEB-1991: 91US-0662276.

XX 13-DEC-1991: 91US-0807529.

XX 25-MAR-1992: 92US-0857311.

XX 15-MAY-1992: 92US-0884718.

XX 15-JAN-1993: 93US-0006116.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Garman RD, Greenstein JL, Kuo M, Briner TJ, Morville M, Gelfer ML;

XX WPI: 2000-146862/13.

XX Peptides of human T cell reactive feline protein for treating

XX sensitivity to cat protein allergens comprise at least one T cell

XX epitope recognized by a T cell receptor specific for the human T cell

XX reactive feline protein -

XX Example 5; Column 93-94; 105pp; English.

XX This invention describes a novel peptide (I) of human T cell reactive

XX feline protein (hTRFP) having at least one T cell epitope recognized

XX by a T cell receptor specific for the human T cell reactive feline

XX protein, the peptide consisting of at least 7-30 amino acids, and having

XX an amino acid sequence derived from an amino acid sequence comprising 94,

XX 96, 97, 109, or 111 residues, given in the specification. The peptides

XX down regulate the immune response to the allergen. The peptides have

XX reduced immunoglobulin E binding and reduce T cell responsiveness. The

XX peptide (I) is useful in compositions for treating sensitivity to a cat

XX protein allergen in a subject. This sequence represents the human TRFP

XX derived peptide used in the method of the invention.

XX Sequence 30 AA;

Query Match 16.2%; Score 18; DB 21; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.7e-10; Mismatches 0; Indels 0; Gaps 0;

QY 83 MTTTSSSKDCMGAVONT 100

Db 1 mtltsskdcmgavqnt 18

RESULT 81

AAY25534

ID AAY25534 standard; peptide: 17 AA.

XX AAY25534;

XX 30-SEP-1999 (first entry)

XX Human MHC Class II desensitizing peptide #8.

XX Major histocompatibility complex; class II; desensitizing; human;  
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.  
XX  
OS Synthetic.  
XX  
PN MO9934826-A1.  
PD 15-JUL-1999.  
XX  
PF 11-JAN-1999; 99MO-GB00080.  
XX  
PR 21-SEP-1998; 98GB-0020474.  
XX  
PR 09-JAN-1998; 98GB-0000445.  
XX  
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
XX  
PI Kay AB, Larche M;  
XX  
DR WPI; 1999-458255/38.  
XX  
PT Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX  
CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting mollusc larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC tenbrion mollusc beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a peptide used to desensitize the human major  
CC histocompatibility complex (MHC) class II response to the Fel d1 Chain 2  
CC allergen.  
XX  
SQ Sequence 17 AA:  
XX  
Query Match 15.3%; Score 17; DB 20; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 26 CPTFYDVFPFVANGNEL 42  
DB 1 fydvfavfavnngnell 17  
XX  
RESULT 82  
AAB28948  
ID AAB28948 standard; Peptide; 17 AA.  
XX  
AC AAB28948;  
XX  
DT 29-JAN-2001 (first entry)  
XX  
DE Peptide Fel 33.  
XX  
KW Cat; allergy; human T cell reactive feline protein; hTRFP;  
KW immunotherapy.  
XX

OS Fells sp.  
XX  
PN US6120769-A.  
XX  
PD 19-SEP-2000.  
XX  
PF 28-APR-1995; 95US-0431184.  
XX  
PR 02-SEP-1994; 94US-0300928.  
XX  
PR 03-NOV-1989; 89US-0431565.  
XX  
PR 28-FEB-1991; 91US-0662276.  
XX  
PR 13-DEC-1991; 91US-0807529.  
XX  
PR 25-MAR-1992; 92US-0857311.  
XX  
PR 15-MAY-1992; 92US-0884718.  
XX  
PR 15-JAN-1993; 93US-0006116.  
XX  
PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
XX  
PI Gefter ML, Garman RD, Greenstein JL, Bond JF;  
XX  
DR WPI; 2000-601477/57.  
XX  
PT Detecting, preventing and treating sensitivity to cat protein allergen  
XX comprises combining a biological sample with a human T cell reactive  
XX feline protein and determining the extent of binding that occurs.  
XX  
PS Example 7; Figure 18; 106pp; English.  
XX  
CC The present invention relates to the detection of sensitivity to a cat  
CC protein allergen by combining a blood sample from a subject with a  
CC peptide of human T cell reactive feline protein (hTRFP). This method  
CC and the hTRFP peptides are useful for diagnosing, preventing and  
CC treating cat allergies by reducing or abolishing an individual's  
CC allergic response to a cat allergen. DNA encoding the TRFP may be  
CC used as probes to locate equivalent sequences present in other species.  
CC These may further be used to study the mechanism of immunotherapy of  
CC cat allergy, and to design modified derivatives, analogues or  
CC functional equivalents useful in immunotherapy. The present  
CC sequence was used in the invention.  
XX  
SQ Sequence 17 AA:  
XX  
Query Match 15.3%; Score 17; DB 21; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 FYDVFPFVANGNELLD 45  
DB 1 fydvfavfavnngnellld 17  
XX  
RESULT 83  
AAB28951  
ID AAB28951 standard; Peptide; 17 AA.  
XX  
AC AAB28951;  
XX  
DT 29-JAN-2001 (first entry)  
XX  
DE Peptide Fel 36.  
XX  
KW Cat; allergy; human T cell reactive feline protein; hTRFP;  
KW immunotherapy.  
XX  
OS Fells sp.  
XX  
PN US6120769-A.  
XX  
PD 19-SEP-2000.  
XX  
PP 28-APR-1995; 95US-0431184.  
XX

PR 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 PI Geffer ML, Garman RD, Greenstein JL, Bond JF;  
 DR WPI; 2000-601477/57.  
 XX  
 PT Detecting, preventing and treating sensitivity to cat protein allergen  
 PT comprises combining a biological sample with a human T cell reactive  
 PT feline protein and determining the extent of binding that occurs  
 PS  
 XX Example 7; Figure 18; 106pp; English.  
 CC The present invention relates to the detection of sensitivity to a cat  
 CC protein allergen by combining a blood sample from a subject with a  
 CC peptide of human T cell reactive feline protein (hTRFP). This method  
 CC and the hTRFP peptides are useful for diagnosing, preventing and  
 CC treating cat allergies by reducing or abolishing an individual's  
 CC allergic response to a cat allergen. DNA encoding the TRFP may be  
 CC used as probes to locate equivalent sequences present in other species.  
 CC These may further be used to study the mechanism of immunotherapy of  
 CC cat allergy, and to design modified derivatives, analogues or  
 CC functional equivalents useful in immunotherapy. The present  
 CC sequence was used in the invention.  
 XX  
 SQ Sequence 17 AA;

Query Match 15.3%; Score 17; DB 21; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 ERTAMKRTQDCYVENG 73  
 ||||||||||||||||  
 Db 1 ertamkrtqdcyvengl 17

## RESULT 84

AY87689  
 ID AAY87689 standard; Protein: 17 AA.  
 XX  
 AC AAY87689;  
 XX  
 DT 22-AUG-2000 (first entry)  
 XX  
 DE Feline human TRFP peptide Fel 33.  
 XX  
 KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
 KW antiallergic; T cell stimulator; diagnostic; immunotherapy.  
 XX  
 OS Felis sp.  
 XX  
 PN US6048962-A.  
 XX  
 PD 11-APR-2000.  
 XX  
 PF 27-APR-1995; 95US-0430014.  
 XX  
 PR 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX

PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Kuo M, Rogers BL, Geffer ML, Morgenstern JP, Brauer AW;  
 PI Greenstein JL, Griffith J, Garman RD;  
 XX  
 DR WPI; 2000-316905/27.  
 XX  
 PT New human T cell reactive feline protein useful for reducing or  
 PT abolishing individual's allergic response to cat allergen comprising  
 PT two different covalently linked peptide chains  
 PS  
 XX Example 5; Column 89-90; 106pp; English.

CC This invention describes a novel naturally occurring cat protein allergen  
 CC (1), human T cell reactive feline protein (TRFP), comprising two  
 CC different covalently linked peptide chains with a molecular weight of 20  
 CC kD, 40 kD or 130 kD under non-reducing conditions and 5 kD or 10-18 kD  
 CC under reducing conditions. The products of the invention have  
 CC antiallergic activity and act as human T cell stimulators. TRFP is useful  
 CC for reducing or preventing the adverse effects of cat allergens on cat  
 CC allergic individuals and in ex vivo diagnostic tests to determine which  
 CC peptides cause sensitivity so as to selectively use them to desensitize  
 CC a cat sensitive individual. Purified TRFP is also useful for studying  
 CC the mechanism of immunotherapy of cat allergy and to design modified  
 CC derivatives, analogs or functional equivalents that are more useful in  
 CC immunotherapy against cat allergy. DNA sequences encoding TRFP are  
 CC useful as probes to locate equivalent sequences present in other species  
 CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics  
 CC and/or therapeutics. Fully defined and characterized TRFP provides  
 CC complete and a very simple desensitization therapy. This sequence  
 CC represents a human T cell reactive feline protein (also known as Fel d I)  
 CC derived peptide Fel 33 which is described in the method of the invention.  
 XX  
 SQ Sequence 17 AA;

Query Match 15.3%; Score 17; DB 21; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 FYDVFPAVANGNELLD 45  
 ||||||||||||||||  
 Db 1 fydvfpavangnellld 17

## RESULT 85

AY87692  
 ID AAY87692 standard; Protein: 17 AA.  
 XX  
 AC AAY87692;  
 XX  
 DT 22-AUG-2000 (first entry)  
 XX  
 DE Feline human TRFP peptide Fel 36.  
 XX  
 KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
 KW antiallergic; T cell stimulator; diagnostic; immunotherapy.  
 XX  
 OS Felis sp.  
 XX  
 PN US6048962-A.  
 XX  
 PD 11-APR-2000.  
 XX  
 PF 27-APR-1995; 95US-0430014.  
 XX  
 PR 02-SEP-1994; 94US-0300928.  
 PR 03-NOV-1989; 89US-0431565.  
 PR 28-FEB-1991; 91US-0662276.  
 PR 13-DEC-1991; 91US-0807529.  
 PR 25-MAR-1992; 92US-0857311.  
 PR 15-MAY-1992; 92US-0884718.  
 PR 15-JAN-1993; 93US-0006116.  
 XX

```

XX (IMMU-) IMMUNOLOGIC PHARM CORP.
PA Kuo M, Rogers BL, Morgenstern JP, Brauer AW;
PI Greenstein JL, Griffith IJ, Garman RD;
XX WPI: 2000-316905/27.
XX
PT New human T cell reactive feline protein useful for reducing or
PT abolishing individual's allergic response to cat allergen comprising
XX two different covalently linked peptide chains
XX
PS Example 5: Column 91-92; 106pp; English.
XX
CC This invention describes a novel naturally occurring cat protein allergen
CC (I), human T cell reactive feline protein (TRFP), comprising two
CC different covalently linked peptide chains with a molecular weight of 20
CC kD, 40 kD or 130 kD under reducing conditions and 5 kD or 10-18 kD
CC under reducing conditions. The products of the invention have
CC antiallergic activity and act as human T cell stimulators. TRFP is useful
CC for reducing or preventing the adverse effects of cat allergens on cat
CC allergic individuals and in ex vivo diagnostic tests to determine which
CC peptides cause sensitivity to cat allergy and to design modified
CC a cat sensitive individual's DNA sequences encoding TRFP are
CC the mechanism of immunotherapy of cat allergy and to design modified
CC immunotherapy against cat allergy. DNA sequences encoding TRFP are
CC useful as probes to locate equine, canine, feline, and other species
CC (goats, sheep, dogs, rabbits or horses) that may be useful in diagnostics
CC and/or therapeutics. Fully defined and characterized TRFP provides
CC complete and a very simple desensitization therapy. This sequence
CC represents a human T cell reactive feline protein (also known as Fel d 1)
CC derived peptide Fel 36 which is described in the method of the invention.
XX
SQ Sequence 17 AA:

Query Match 15.3%; Score 17; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 ERTAMKRIQCYVENG 73
DB 1 ertamkriqcyvengl 17

RESULT 86
AA90119
ID AA90119 standard; Peptide: 17 AA.
XX
AC AA90119;
XX
DE 13-JUL-2000 (first entry)
XX
DE Cat TRFP derived peptide, peptide Fel 33.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PE 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.

```

```

PR 15-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Morgenstern JP, Griffith IJ, Rogers BL;
XX WPI: 2000-181812/16.
XX
PT New human T cell reactive feline protein, useful for desensitizing cat
PT allergic individuals to cat allergens
XX
PS Example 7: Fig 18; 108pp; English.
XX
CC This sequence is a peptide derived from the human T cell reactive feline
CC protein (TRFP) of the invention. The protein is a cat protein allergen,
CC and was isolated from a vacuum bag extract obtained by affinity
CC purification of house dust collected from several homes with cats. TRFP
CC is composed of two covalently linked peptide chains, and is also referred
CC to as Fel d I. TRFP and its peptides are useful for reducing or
CC preventing the adverse effects that exposure to cat allergens normally
CC has on cat allergic individuals (i.e. to desensitize individuals to cat
CC allergens or block the effect of the allergens). TRFP is also used in
CC methods of diagnosing sensitivity to Felis domesticus in an individual.
CC DNA sequences encoding TRFP can be used as probes to locate equivalent
CC sequences present in other species, e.g. goat, sheep, horse, rabbit and
CC dog, that may be useful in diagnostic and/or therapeutic applications.
XX
SQ Sequence 17 AA:

Query Match 15.3%; Score 17; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 FYDVFPAVANGNELLD 45
DB 1 fydvfpavangnellld 17

RESULT 87
AA90122
ID AA90122 standard; Peptide: 17 AA.
XX
AC AA90122;
XX
DE 13-JUL-2000 (first entry)
XX
DE Cat TRFP derived peptide, peptide Fel 36.
XX
KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;
KW diagnosis; goat; sheep; horse; rabbit; dog.
XX
OS Felis domesticus.
XX
PN US6025162-A.
XX
PD 15-FEB-2000.
XX
PE 28-APR-1995; 95US-0430944.
XX
PR 02-SEP-1994; 94US-0300928.
PR 03-NOV-1989; 89US-0431565.
PR 28-FEB-1991; 91US-0662276.
PR 13-DEC-1991; 91US-0807529.
PR 25-MAR-1992; 92US-0857311.
PR 13-MAY-1992; 92US-0884718.
PR 15-JAN-1993; 93US-0006116.
XX
XX (IMMU-) IMMUNOLOGIC PHARM CORP.
XX
XX Morgenstern JP, Griffith IJ, Rogers BL;

```

Query Match	15.3%;	Score 17;	DB 21;	Length 17;
Best Local Similarity	100.0%;	Pred. No. 1.4e-09;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT 88  
AAV51486  
ID AAV51486 standard; Protein; 17 AA.

DT 22-MAY-2000 (first entry)

Human TRFP derived peptide Fel 33.

KW T-cell reactive feline protein; TR

KW sensitivity; cat protein allergen; human.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 13-DEC-1991; 91US-0807529.

PR 15-MAY-1992; 92US-0884718.

[illegible]

XX

XX  
XX  
END

33 XX

PT sensitivity to cat protein allergens comprise at least one T

PT reactive feline protein -

CC This invention describes a novel peptide (I) of human T cell reactive  
CC feline protein (hTRFP) having at least one T cell epitope recognized  
CC by a T cell receptor specific for the human T cell reactive feline  
CC protein, the peptide consisting of at least 7-30 amino acids, and having  
CC an amino acid sequence derived from an amino acid sequence comprising 94  
CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
CC down regulate the immune response to the allergen. The peptides have  
CC reduced immunoglobulin E binding and reduce T cell responsiveness. The  
CC peptide (I) is useful in compositions for treating sensitivity to a cat  
CC protein allergen in a subject. This sequence represents the human TRFP  
CC derived peptide used in the method of the invention.

**SQ Sequence 17 AA;**

Query Match	15.3%;	Score 17;	DB 21;	Length 17;
Best Local Similarity	100.0%;	Pred. No. 1.4e-09;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT 89  
AAV51489  
ID AAV51489 standard; Protein; 17 AA

AC AAY51489;

DT 22-MAY-2000 (first entry)

DE Human TRFP derived peptide Fel 36.

KW T-cell reactive feline protein; TRFP: T cell epitope; T cell receptor;

KW sensitivity; cat protein allergen; human.

OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

PR 03-NOV-1989; 89US-0431565.

PR 13-DEC-1991; 91US-0807529.

PR 15-MAY-1992; 92US-0884718.

XX

XX

XX

XX

PT sensitivity to cat p

PT reactive feline protein -

PS Example 5; Column 91-92; 105pp; English

CC This invention describes a novel peptide (I) of human T cell reactive

CC by a T cell receptor specific for the human T cell reactive feline

CC an amino acid sequence derived from an amino acid sequence comprising 94,  
 CC 96, 97, 109, or 111 residues, given in the specification. The peptides  
 CC down regulate the immune response to the allergen. The peptides have  
 CC reduced immunoglobulin binding and reduce T cell responsiveness. The  
 CC peptide (1) is useful in compositions for treating sensitivity to a cat  
 CC protein allergen in a subject. This sequence represents the human TRFP  
 CC derived peptide used in the method of the invention.  
 XX  
 SO Sequence 17 AA:

Query Match 15.3% Score 17; DB 21; Length 17;  
 Best Local Similarity 100.0%; Pred No. 1 4e-09;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ERTAKKRIQDCYENG 73  
 |||||  
 DB 1 ertankkriqdcyengl 17

## RESULT 90

AAW40943  
 ID AAW40943 standard; peptide; 16 AA.

AC AAW40943;  
 DT 09-APR-1998 (first entry)

DE Cryptic peptide 2.1 of chain 2 of major cat allergen Fel d1.

KW Major cat allergen; Fel d1; chain 1; cryptic peptide; T-cell; asthma;  
 KM atopic allergy; therapy.

OS Fells sp.

PN WO9735193-A1.

PD 25-SEP-1997.

PF 20-MAR-1997; 97WO-GB00783.

PR 24-APR-1996; 96GB-0008430.

PK 21-MAR-1996; 96GB-0005904.

PA (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.

PI Kay AB, Larche M;

DR WPI: 1997-480354/44.

CC Determining if peptide of protein is cryptic peptide - by comparing  
 CC its reactivity with pre-challenged and non-pre-challenged T cells,  
 CC useful to diagnose or treat atopic condition, e.g. asthma

Claim 12; Fig 2; 49pp; English.

CC This sequence represents a cryptic peptide of the chain 2 sequence of the  
 CC major cat allergen Fel d1. This sequence can be used in the method of the  
 CC invention. The method of the invention is for determining if a peptide of  
 CC a protein is a cryptic peptide, and comprises: (a) exposing T-cells to  
 CC the peptide in a primary challenge, and measuring the reactivity of the  
 CC T-cells to the peptide; (b) exposing pre-challenged T-cells, obtained by  
 CC exposure to the protein, to the peptide in a secondary challenge, and  
 CC measuring the reactivity of the pre-challenged T-cells to the peptide;  
 CC and (c) determining the reactivity of the pre-challenged T-cells to the peptide;  
 CC reactivity is observed in the secondary, but not in the primary  
 CC challenge. Peptides identified as being cryptic can be used to diagnose  
 CC or treat an atopic allergy, e.g. asthma. They can also be used in to  
 CC screen compounds for therapeutic activity, e.g. asthma therapeutic  
 CC activity.

XX Sequence 16 AA:

Query Match 14.4% Score 16; DB 18; Length 16;  
 Best Local Similarity 100.0%; Pred No. 1 3e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 VKNAETCPIFYDVFA 35  
 |||||  
 DB 1 vknaetcpifydvffa 16

## RESULT 91

AAW40948  
 ID AAW40948 standard; peptide; 16 AA.

AC AAW40948;  
 DT 09-APR-1998 (first entry)

DE Cryptic peptide 2.6 of chain 2 of major cat allergen Fel d1.

KW Major cat allergen; Fel d1; chain 1; cryptic peptide; T-cell; asthma;  
 KM atopic allergy; therapy.

OS Fells sp.

PN WO9735193-A1.

PD 25-SEP-1997.

PF 20-MAR-1997; 97WO-GB00783.

PR 24-APR-1996; 96GB-0008430.

PK 21-MAR-1996; 96GB-0005904.

PA (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.

PI Kay AB, Larche M;

DR WPI: 1997-480354/44.

CC Determining if peptide of protein is cryptic peptide - by comparing  
 CC its reactivity with pre-challenged and non-pre-challenged T cells,  
 CC useful to diagnose or treat atopic condition, e.g. asthma

Claim 12; Fig 2; 49pp; English.

CC This sequence represents a cryptic peptide of the chain 2 sequence of the  
 CC major cat allergen Fel d1. This sequence can be used in the method of the  
 CC invention. The method of the invention is for determining if a peptide of  
 CC a protein is a cryptic peptide, and comprises: (a) exposing T-cells to  
 CC the peptide in a primary challenge, and measuring the reactivity of the  
 CC T-cells to the peptide; (b) exposing pre-challenged T-cells, obtained by  
 CC exposure to the protein, to the peptide in a secondary challenge, and  
 CC measuring the reactivity of the pre-challenged T-cells to the peptide;  
 CC and (c) determining the reactivity of the pre-challenged T-cells to the peptide;  
 CC reactivity is observed in the secondary, but not in the primary  
 CC challenge. Peptides identified as being cryptic can be used to diagnose  
 CC or treat an atopic allergy, e.g. asthma. They can also be used in to  
 CC screen compounds for therapeutic activity, e.g. asthma therapeutic  
 CC activity.

XX Sequence 16 AA:

Query Match 14.4% Score 16; DB 18; Length 16;  
 Best Local Similarity 100.0%; Pred No. 1 3e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 RTAKKRIQDCYENG 73  
 |||||  
 DB 1 rtankkriqdcyengl 16



RESULT 92  
 AAM40949 standard; peptide; 16 AA.  
 XX  
 AC AAM40949;  
 XX  
 DT 09-APR-1998 (first entry)  
 XX  
 DE Cryptic peptide 2.7 of chain 2 of major cat allergen Fel di.  
 XX  
 KW Major cat allergen; Fel di; chain 1; cryptic peptide; T-cell; asthma;  
 XX atopic allergy; therapy.  
 XX  
 OS Felis sp.  
 XX  
 PN WO9735193-A1.  
 PD 25-SEP-1997.  
 XX  
 PF 20-MAR-1997; 97WO-GB00783.  
 XX  
 PR 24-APR-1996; 96GB-0008430.  
 XX 21-MAR-1996; 96GB-0005904.  
 XX  
 PA (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 XX  
 PI Kay AB, Larche M;  
 XX  
 DR WPI; 1997-480354/44.  
 XX  
 PT Determining if peptide of protein is cryptic peptide - by comparing  
 XX its reactivity with pre-challenged and non-pre-challenged T cells,  
 PT useful to diagnose or treat atopic condition, e.g. asthma  
 XX  
 PS Claim 12; Fig 2; 49pp; English.  
 XX  
 CC This sequence represents a cryptic peptide of the chain 2 sequence of the  
 CC major cat allergen Fel di. This sequence can be used in the method of the  
 CC invention. The method of the invention is for determining if a peptide of  
 CC a protein is a cryptic peptide, and comprises: (a) exposing T-cells to  
 CC the peptide in a primary challenge, and measuring the reactivity of the  
 CC T-cells to the peptide; (b) exposing pre-challenged T-cells, obtained by  
 CC exposure to the protein, to the peptide in a secondary challenge, and  
 CC measuring the reactivity of the pre-challenged T-cells to the peptide;  
 CC and (c) determining the peptide to be a cryptic peptide if T-cell  
 CC reactivity is observed in the secondary, but not in the primary  
 CC challenge. Peptides identified as being cryptic can be used to diagnose  
 CC or treat an atopic allergy, e.g. asthma. They can also be used in to  
 CC screen compounds for therapeutic activity, e.g. asthma therapeutic  
 CC activity.  
 CC  
 SQ Sequence 16 AA:  
 XX  
 XX  
 Query Match 14.4%; Score 16; DB 18; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 67 CYVENGILSRVLDGLV 82  
 DB 1 cyvengilsrvldgvlv 16  
 XX  
 RESULT 93  
 AAM40950 standard; peptide; 16 AA.  
 XX  
 AC AAM40950;  
 XX  
 DT 09-APR-1998 (first entry)  
 XX  
 DE Cryptic peptide 2.8 of chain 2 of major cat allergen Fel di.  
 XX

KW Major cat allergen; Fel di; chain 1; cryptic peptide; T-cell; asthma;  
 KW atopic allergy; therapy.  
 XX  
 OS Felis sp.  
 XX  
 PN WO9735193-A1.  
 PD 25-SEP-1997.  
 XX  
 PF 20-MAR-1997; 97WO-GB00783.  
 XX  
 PR 24-APR-1996; 96GB-0008430.  
 XX 21-MAR-1996; 96GB-0005904.  
 XX  
 PA (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 XX  
 PI Kay AB, Larche M;  
 XX  
 DR WPI; 1997-480354/44.  
 XX  
 PT Determining if peptide of protein is cryptic peptide - by comparing  
 XX its reactivity with pre-challenged and non-pre-challenged T cells,  
 PT useful to diagnose or treat atopic condition, e.g. asthma  
 XX  
 PS Claim 12; Fig 2; 49pp; English.  
 XX  
 CC This sequence represents a cryptic peptide of the chain 2 sequence of the  
 CC major cat allergen Fel di. This sequence can be used in the method of the  
 CC invention. The method of the invention is for determining if a peptide of  
 CC a protein is a cryptic peptide, and comprises: (a) exposing T-cells to  
 CC the peptide in a primary challenge, and measuring the reactivity of the  
 CC T-cells to the peptide; (b) exposing pre-challenged T-cells, obtained by  
 CC exposure to the protein, to the peptide in a secondary challenge, and  
 CC measuring the reactivity of the pre-challenged T-cells to the peptide;  
 CC and (c) determining the peptide to be a cryptic peptide if T-cell  
 CC reactivity is observed in the secondary, but not in the primary  
 CC challenge. Peptides identified as being cryptic can be used to diagnose  
 CC or treat an atopic allergy, e.g. asthma. They can also be used in to  
 CC screen compounds for therapeutic activity, e.g. asthma therapeutic  
 CC activity.  
 CC  
 SQ Sequence 16 AA:  
 XX  
 XX  
 Query Match 14.4%; Score 16; DB 18; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 74 ISRVLDGLVWTTIRSS 89  
 DB 1 isrvldglvmttirs 16  
 XX  
 RESULT 94  
 AAM40952 standard; peptide; 16 AA.  
 XX  
 AC AAM40952;  
 XX  
 DT 09-APR-1998 (first entry)  
 XX  
 DE Cryptic peptide 2.10 of chain 2 of major cat allergen Fel di.  
 XX  
 KW Major cat allergen; Fel di; chain 1; cryptic peptide; T-cell; asthma;  
 KW atopic allergy; therapy.  
 XX  
 OS Felis sp.  
 XX  
 PN WO9735193-A1.  
 PD 25-SEP-1997.  
 XX  
 PF 20-MAR-1997; 97WO-GB00783.  
 XX



PT Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX  
CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenebrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a peptide used to desensitize the human major  
CC histocompatibility complex (MHC) class II response to the Fel d1 Chain 2  
CC allergen.  
XX  
SQ Sequence 16 AA;  
XX  
XX  
Query Match 14.4%; Score 16; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 48 LTKYNATEPERTAMKK 63  
XXXXXXXXXXXXXXXXXXXX  
Db 1 ltkynatepertamkk 16  
XX  
RESULT 97  
AAV25537  
ID AAV25537 standard; peptide; 16 AA.  
XX  
AC AAV25537;  
XX  
DT 30-SEP-1999 (first entry)  
XX  
DE Human MHC Class II desensitizing peptide #11.  
XX  
KW Major histocompatibility complex; class II; desensitizing; human;  
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.  
XX  
OS Synthetic.  
XX  
PN WO9934826-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 11-JAN-1999; 99WO-GB00080.  
XX  
PR 21-SEP-1998; 98GB-0020474.  
PR 09-JAN-1998; 98GB-0000445.  
XX  
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
XX  
PI Kay AB, Larche M;  
XX  
DR WPI; 1999-458255/38.  
XX  
PT Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX

CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC Tenebrio molitor beetle, mammals such as cat, dog, horse, cow, pig,  
CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a peptide used to desensitize the human major  
CC histocompatibility complex (MHC) class II response to the Fel d1 Chain 2  
CC allergen.  
XX  
SQ Sequence 16 AA;  
XX  
XX  
Query Match 14.4%; Score 16; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 59 TAMKKIDDCYENGCI 74  
XXXXXXXXXXXXXXXXXXXX  
Db 1 tamkkidcyengcli 16  
XX  
RESULT 98  
AAV25538  
ID AAV25538 standard; peptide; 16 AA.  
XX  
AC AAV25538;  
XX  
DT 30-SEP-1999 (first entry)  
XX  
DE Human MHC Class II desensitizing peptide #12.  
XX  
KW Major histocompatibility complex; class II; desensitizing; human;  
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;  
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;  
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.  
XX  
OS Synthetic.  
XX  
PN WO9934826-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 11-JAN-1999; 99WO-GB00080.  
XX  
PR 21-SEP-1998; 98GB-0020474.  
PR 09-JAN-1998; 98GB-0000445.  
XX  
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
XX  
PI Kay AB, Larche M;  
XX  
DR WPI; 1999-458255/38.  
XX  
PT Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX  
CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the

peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tentorio mollitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a peptide used to desensitize the human major histocompatibility complex (MHC) class II response to the Fel d1 Chain 2 allergen.

Sequence 16 AA:

Query Match 14.4%; Score 16; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVNGLSRYLQIV 82  
Db 1 CYVNGLSRYLQIV 16

RESULT 99  
AAV25539  
ID AAV25539 standard; peptide; 16 AA.  
AC AAV25539;  
XX  
DT 30-SEP-1999 (first entry)  
XX  
DE Human MHC Class II desensitizing peptide #13.  
XX  
KW Major histocompatibility complex; class II; desensitizing; human;  
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KW screw worm fly; grain weevil; silkworm; bee moth; mealworm; cat;  
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.  
XX  
OS Synthetic.  
XX  
PN WO9934826-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 11-JAN-1999; 99WO-GB00080.  
XX  
PR 21-SEP-1998; 98GB-0020474.  
XX  
PR 09-JAN-1998; 98GB-0000445.  
XX  
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
XX  
PI Kay AB, Larche M;  
XX  
PT WPI; 1999-458255/38.  
XX  
DR Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX  
CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,

the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tentorio mollitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a peptide used to desensitize the human major histocompatibility complex (MHC) class II response to the Fel d1 Chain 2 allergen.

Sequence 16 AA:

Query Match 14.4%; Score 16; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 SRVLDGIWMTTSSK 90  
Db 1 SRVLDGIWMTTSSK 16

RESULT 100  
AAV25540  
ID AAV25540 standard; peptide; 16 AA.  
AC AAV25540;  
XX  
DT 30-SEP-1999 (first entry)  
XX  
DE Human MHC Class II desensitizing peptide #14.  
XX  
KW Major histocompatibility complex; class II; desensitizing; human;  
KW allergen; grass; tree; weed; pollen; fungi; mould; food; insect; stinging;  
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;  
KW screw worm fly; grain weevil; silkworm; bee moth; mealworm; cat;  
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;  
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.  
XX  
OS Synthetic.  
XX  
PN WO9934826-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 11-JAN-1999; 99WO-GB00080.  
XX  
PR 21-SEP-1998; 98GB-0020474.  
XX  
PR 09-JAN-1998; 98GB-0000445.  
XX  
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
XX  
PI Kay AB, Larche M;  
XX  
PT WPI; 1999-458255/38.  
XX  
DR Desensitizing patients to polypeptide allergens  
XX  
PS Claim 7; Figure 9; 117pp; English.  
XX  
CC This invention describes a novel method of desensitizing a patient to a  
CC polypeptide allergen and comprises administering to the patient a peptide  
CC derived from the allergen where restriction to a MHC Class II molecule  
CC possessed by the patient can be demonstrated for the peptide and the  
CC peptide is able to induce a late phase response in an individual who  
CC possesses the MHC Class II molecule. The methods can be used for  
CC desensitizing patients to allergens present in e.g. grass, tree and weed  
CC (including ragweed) pollens, fungi and moulds, foods, stinging insects,  
CC the chironomidae (non-biting midges), spiders and mites, housefly, fruit  
CC fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee,  
CC non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of  
CC tentorio mollitor beetle, mammals such as cat, dog, horse, cow, pig,

CC sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to  
CC produce immunological vaccines which may be used to prevent and/or treat  
CC conditions involving hypersensitivity to allergens. This sequence  
CC represents a peptide used to desensitise the human major  
CC histocompatibility complex (MHC) class II response to the Fel d1 Chain 2  
CC allergen.

XX

SQ Sequence 16 AA;

Query Match 14.4%; Score 16; DB 20; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ISSSKDCMGEAVONTV 101

|||||

Db 1 isskdcmgaeavqntv 16

Search completed: July 9, 2002, 16:19:12  
Job time: 62 sec

---

**THIS PAGE BLANK (USPTO)**

1

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:18:10 ; Search time 16.02 Seconds

(Without alignments)  
665.788 Million cell updates/sec

Title: US-09-662-784-6

Perfect score: 111

Sequence: 1 DTRGALVIALVLTQALGV.....CMGEAVONTVEDLKINTLGR 111

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_71:\*\*

1: pirl:\*\*  
2: pirl:\*\*  
3: pirl:\*\*  
4: pirl:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	109	98.2	109	2	C56413	major allergen Fel
2	81	73.0	107	2	JC1127	major allergen cha
3	8	7.2	108	2	S40149	integrin alpha-7C
4	8	7.2	129	1	I61187	alpha-7A integrin
5	8	7.2	139	1	C45051	lamprin 2 precursor
6	8	7.2	148	2	S40148	integrin alpha-7A
7	8	7.2	167	2	AH1106	hypothetical prote
8	8	7.2	193	2	AH3639	cytochrome b561 (1
9	8	7.2	299	2	T06960	probable membrane
10	8	7.2	378	2	JC5649	N-acetylglucosamin
11	8	7.2	378	2	E82254	probable MFS trans
12	8	7.2	401	2	E82951	integrin alpha cha
13	8	7.2	1106	2	S38783	integrin alpha cha
14	8	7.2	1135	2	I61186	alpha-7 integrin -
15	8	7.2	1137	2	JC5950	integrin alpha-7 c
16	8	7.2	20	2	B53283	major cat allergen
17	7	6.3	73	2	T20911	hypothetical prote
18	7	6.3	77	2	D83680	hypothetical prote
19	7	6.3	87	2	G46449	hypothetical prote
20	7	6.3	103	2	G75257	hypothetical prote
21	7	6.3	137	2	AEO751	probable lipoprote
22	7	6.3	198	2	G59094	hypothetical prote
23	7	6.3	203	2	B75497	hypothetical prote
24	7	6.3	244	2	D75505	hypothetical prote
25	7	6.3	245	2	A97381	hypothetical trans
26	7	6.3	245	2	AH2588	transcription regu
27	7	6.3	250	2	T50677	bacteriorhodopsin
28	7	6.3	253	2	T46454	hypothetical prote
29	7	6.3	262	2	AH1601	phosphatidate cyti

30	7	6.3	266	2	T36287	probable integral
31	7	6.3	268	2	A84746	probable NM (no a
32	7	6.3	279	2	S75087	hypothetical prote
33	7	6.3	303	2	T00479	probable phosphati
34	7	6.3	321	1	E69687	do-type ubiquinol
35	7	6.3	322	2	B83579	hypothetical prote
36	7	6.3	337	1	G64125	hemlin transport pr
37	7	6.3	341	2	C72403	hypothetical prote
38	7	6.3	350	2	S75065	sensory transducti
39	7	6.3	355	2	JC1249	peroxidase (EC 1.1
40	7	6.3	360	2	E95126	amino transferase,
41	7	6.3	364	2	S34355	peroxidase (EC 1.1
42	7	6.3	371	2	A97997	pyridoxal-phosphat
43	7	6.3	374	2	B82488	hypothetical prote
44	7	6.3	385	2	E83506	probable MFS trans
45	7	6.3	390	2	H70904	probable lprk prot

#### ALIGNMENTS

RESULT 1  
C56413  
major allergen Fel di chain 2 precursor - cat  
C:Species: Felis silvestris catus (domestic cat)  
C:Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 01-Dec-2000  
C:Accession: C56413; JC1145  
R:Morgenstern, J.P.; Griffith, I.J.; Brauer, A.W.; Rogers, B.L.; Bond, J.F.; Chapman, Proc. Natl. Acad. Sci. U.S.A. 88, 9690-9694, 1991  
A:Title: Amino acid sequence of Fel di, the major allergen of the domestic cat: prot  
A:Reference number: A56413; MUID:92052157  
A:Accession: C56413  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-109 <MOR>  
A:Cross-references: GB:M77341; NID:g163822; PIDN:AAC41616.1; PID:g163823  
R:Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L. Gene 113, 263-268, 1992  
A:Title: Expression and genomic structure of the genes encoding Fdi, the major allergy  
A:Reference number: JC1126; MUID:92241678  
A:Accession: JC1145  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 18-109 <GR1>  
A:Experimental source: salivary gland  
C:Keywords: glycoprotein  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:50/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 98.2% Score 109; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 1e-100;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 MRGALLVIALVLTQALGVKAETCPFYDVFPAVANGNELLDLSTKVNATEPERAMK 62  
DB 1 MRGALLVIALVLTQALGVKAETCPFYDVFPAVANGNELLDLSTKVNATEPERAMK 60  
OY 63 KIQDCYVENGLISRVLDGLVMTTSSSKDCMGEAVONTVEDLKINTLGR 111  
DB 61 KIQDCYVENGLISRVLDGLVMTTSSSKDCMGEAVONTVEDLKINTLGR 109

RESULT 2  
JC1127  
major allergen chain 2 precursor, short form - cat  
C:Species: Felis silvestris catus (domestic cat)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 05-Nov-1999  
C:Accession: JC1127  
R:Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L. Gene 113, 263-268, 1992  
A:Title: Expression and genomic structure of the genes encoding Fdi, the major allergy  
A:Reference number: JC1126; MUID:92241678

A:Accession: JCI127  
 A:Molecule type: DNA  
 A:Residues: 1-107 <GR>  
 A:Cross-references: GB:X62478; NID:9395406; PIDN:CAA44345.1; PID:g395407  
 A:Experimental source: skin  
 C:Genetics:  
 A:Gene: Ch2  
 A:Introns: 21/1, 81/3  
 C:Keywords: glycoprotein  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-107/Product: major allergen chain 2, short form #status predicted <MAT>  
 F:50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 73.0%; Score 81; DB 2; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 6, 5e-73;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 MRGALIVLALVTQALGVKAEFCPIFYDVPFVAVANGNELLDLSITKYNAEPEPTAMK 62  
 Db 1 MRGALIVLALVTQALGVKAEFCPIFYDVPFVAVANGNELLDLSITKYNAEPTAMK 60  
 Oy 63 KIQDCVENGILSRVLDGLVM 83  
 Db 61 KIQDCVENGILSRVLDGLVM 81

RESULT 3  
 S40149  
 Integrin alpha-7C chain - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 29-Sep-1999  
 C:Accession: S40149  
 R:Song, W.K.; Wang, W.; Sato, H.; Bjelser, D.; Kaufman, S.  
 A:Submitted to the EMBL data library, July 1993  
 A:Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal muscle phosphatases.  
 A:Reference number: S40147  
 A:Accession: S40149  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-108 <SON>  
 A:Cross-references: EMBL:X74294; NID:g437912; PIDN:CAA52347.1; PID:g437913  
 C:Superfamily: Integrin alpha-2b chain

Query Match 7.2%; Score 8; DB 2; Length 108;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 LVLVALLV 14  
 Db 79 LVLVALLV 86

RESULT 4  
 I61187  
 alpha-7A integrin - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C>Date: 02-Jul-1995 #sequence\_revision 02-Jul-1996 #text\_change 29-Sep-1999  
 C:Accession: I61187  
 R:Zlobner, B.L.; Vu, M.P.; Walsh, N.; Crawford, J.; Lin, C.S.; Kramer, R.H.  
 J. Biol. Chem. 268, 26773-26783, 1993  
 A:Title: Alternative extracellular and cytoplasmic domains of the integrin alpha 7 subunit  
 A:Reference number: A49691; MUID:94075378  
 A:Accession: I61187  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-129 <RES>  
 A:Cross-references: GB:I23421; NID:g431417; PIDN:AAA16598.1; PID:g450648  
 C:Superfamily: Integrin alpha-2b chain

Query Match 7.2%; Score 8; DB 2; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 LVLVALLV 14  
 Db 61 LVLVALLV 68

RESULT 5  
 C45051  
 Lamprin 2 precursor, long splice form - sea lamprey  
 N:Alternate names: lamprin L-0.9-12  
 N:Contents: lamprin 2 precursor, short splice form (lamprin L-0.9-10)  
 C:Species: Petromyzon marinus (sea lamprey)  
 C>Date: 30-Apr-1993 #sequence\_revision 30-Apr-1993 #text\_change 22-Jun-1999  
 C:Accession: C45051; A45051  
 R:Robson, P.; Wright, G.M.; Slater, E.; Maitl, A.; Rawat, M.; Youson, J.H.; Keeley, F.  
 J. Biol. Chem. 268, 1440-1447, 1993  
 A:Title: Characterization of lamprin, an unusual matrix protein from lamprey cartilage  
 A:Reference number: A45051; MUID:93123269  
 A:Accession: C45051  
 A:Molecule type: mRNA  
 A:Residues: 1-139 <ROB>  
 A:Cross-references: GB:L05925; NID:g213209; PIDN:AAA49269.1; PID:g213210  
 A:Note: sequence extracted from NCBI backbone (NCBIP:122170)  
 A:Accession: A45051  
 A:Molecule type: mRNA  
 A:Residues: 1-85, 105-139 <RO2>  
 A:Cross-references: GB:L05924; NID:g213207; PIDN:AAA49268.1; PID:g213208  
 A:Note: sequence extracted from NCBI backbone (NCBIP:122166)  
 C:Comment: These forms are encoded by a different gene than B45051; we have arbitrary  
 C:Superfamily: lamprin  
 A:Keywords: alternative splicing; cartilage; extracellular matrix  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-139/Product: lamprin 2, long splice form #status predicted <MAT1>  
 F:20-85, 105-139/Product: lamprin 2, short splice form #status predicted <MAT2>

Query Match 7.2%; Score 8; DB 1; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 ALVLALL 13  
 Db 7 ALVLALL 14

RESULT 6  
 S40148  
 Integrin alpha-7A chain - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 29-Sep-1999  
 C:Accession: S40148  
 R:Song, W.K.; Wang, W.; Sato, H.; Bjelser, D.; Kaufman, S.  
 A:Submitted to the EMBL data library, July 1993  
 A:Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal muscle phosphatases.  
 A:Reference number: S40147  
 A:Accession: S40148  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-148 <SON>  
 A:Cross-references: EMBL:X74293; NID:g437910; PIDN:CAA52346.1; PID:g437911  
 C:Superfamily: Integrin alpha-2b chain

Query Match 7.2%; Score 8; DB 2; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 LVLVALLV 14  
 Db 7 LVLVALLV 14



Db 79 LVALLVLT 86

## RESULT 7

AH1106

hypothetical protein lmo0255 [imported] - *Listeria monocytogenes* (strain EGD-e)C:Species: *Listeria monocytogenes*

C&gt;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AH1106

R:Graessner, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Jones, L.M.; Karst, U.

S: Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MIMD:2153729; PMID:11679669

A:Accession: AH1106

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-167 &lt;GTA&gt;

A:Cross-references: GB:NC\_003210; PIRN:CAD00782.1; PIR:q16409620; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0255

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 167;

Matches 8; Conservatively 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALLVLT 15

Db 11 LVALLVLT 18

## RESULT 8

AH3639

cytochrome b561 [imported] - *Brucella melitensis* (strain 16M)C:Species: *Brucella melitensis*

C&gt;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C:Accession: AH3639

R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A:Reference number: AD3252; PMID:11756688

A:Accession: AH3639

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-193 &lt;KUR&gt;

A:Cross-references: GB:AE008918; PIRN:AA154283.1; PIR:q17985259; GSPDB:GN00191

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11041

A:Map position: 11

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 193;

Matches 8; Conservatively 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVALL 13

Db 68 ALLVALL 75

## RESULT 9

T06960

probable membrane protein - *Cyanophora paradoxa* cyanelle

N:Alternate names: hypothetical protein psac region

C:Species: *Cyanophora paradoxa*

C&gt;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 24-Nov-1999

C:Accession: T06960; PS0372

R:Stewart, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohmert, H.J.; Bryant, D.A.

submitted to the EMBL Data Library, July 1995

A:Description: Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*.

A:Reference number: Z15840

A:Accession: T06960

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-299 &lt;STI&gt;

A:Cross-references: EMBL:U30821; NID:q1016083; PIRN:AAA81303.1; PIR:q1016216

A:Experimental source: strain Pringsheim LB555

R:Rhiel, E.; Stewart, V.L.; Gasparich, G.E.; Bryant, D.A.

Gene 112, 123-128, 1992

A:Title: The psac genes of *Synechococcus* sp. PCC7002 and *Cyanophora paradoxa*: cloning

A:Reference number: J50694; MIMD:92201692

A:Accession: PS0372

A:Molecule type: DNA

A:Residues: 138-299 &lt;RHL&gt;

A:Cross-references: GB:M86239; NID:q336634; PIRN:AAA65471.1; PIR:q336637

C:Genetics:

A:Genome: cyanelle

C:Superfamily: conserved hypothetical protein HI0360

C:Keywords: cyanelle; transmembrane protein

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 299;

Matches 8; Conservatively 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALLVLT 15

Db 220 LVALLVLT 227

## RESULT 10

JC5649

N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) - *Vibrio cholerae* (strain 1C:Species: *Vibrio cholerae*

A:Note: non-O1 strain 1148A

C&gt;Date: 28-Oct-1997 #sequence\_revision 28-Oct-1997 #text\_change 05-Nov-1999

C:Accession: JC5649

R:Yamano, N.; Wang, J.; Fujishima, S.

Biochim. Biotechnol. Biochem. 61, 1349-1353, 1997

A:Title: Cloning and sequencing of the genes for N-acetylglucosamine use that constr

A:Reference number: JC5649; MIMD:97446530

A:Accession: JC5649

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-378 &lt;YAM&gt;

A:Cross-references: DDBJ:D87820; NID:q2541899; PIRN:BAA22834.1; PIR:d1023704; PIR:q25

A:Experimental source: strain 1148A

C:Genetics:

A:Gene: nagA

C:Keywords: hydrolase

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 378;

Matches 8; Conservatively 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EAVONTVE 102

Db 312 EAVONTVE 319

## RESULT 11

E82254

N-acetylglucosamine-6-phosphate deacetylase WC0994 [imported] - *Vibrio cholerae* (straC:Species: *Vibrio cholerae*

C&gt;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: E82254

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

Wed Jul 10 08:06:59 2002

us-09-662-784-6\_1.rpr

Page 4

L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833  
A:Accession: E82254  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <HEI>  
A:Cross-references: GB:AE004181; GB:AE003852; NID:95655454; PIDN:AMF94155.1; GSPDB:GN001  
A:Experimental source: serogroup O1, strain N16961; biotype El Tor  
A:Genetics:  
A:Gene: VC0994  
A:Map position: 1

Query Match 7.2% Score 8; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 5, 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 EAVONTVE 102  
|||||||  
Db 312 EAVONTVE 319

RESULT 12  
E82951  
Probable MFS transporter PA5548 [Imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E82951  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Larbig, K.; Lim,  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: E82951  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-401 <STO>  
A:Cross-references: GB:AE004967; GB:AE004091; NID:99951884; PIDN:AG08933.1; GSPDB:GN001  
A:Experimental source: strain PA01  
A:Genetics:  
A:Gene: PA5548

Query Match 7.2% Score 8; DB 2; Length 401;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLALV 14  
|||||||  
Db 281 LTVLALV 288

RESULT 13  
S38783  
Integrin alpha chain - rat (fragment)  
C:Species: *Rattus norvegicus* (Norway rat)  
C:Date: 22-Jan-1994 #sequence\_revision 14-Jul-1995 #text\_change 29-Sep-1999  
C:Accession: S38783; S23600  
R:Kaufman, S.J.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S38783  
A:Accession: S38783  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1106 <KAU>  
A:Cross-references: EMBL:X65036; NID:956392; PIDN:CAA46170.1; PID:956393  
R:Song, W.K.; Wang, W.; Foster, R.F.; Bietzer, D.A.; Kaufman, S.J.  
J. Cell Biol. 117, 643-657, 1992  
A:Title: H36-alpha7 is a novel integrin alpha chain that is developmentally regulated du  
A:Reference number: S23600; MUID:92242309

A:Accession: S23600  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-116, 'R', 118-349, 'D', 351-804, 'R', 806, 'V', 808-1106 <SON>  
A:Cross-references: EMBL:X65036  
C:Superfamily: Integrin alpha-2b chain

Query Match 7.2% Score 8; DB 2; Length 1106;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLALV 14  
|||||||  
Db 1018 LTVLALV 1025

RESULT 14  
I61186  
alpha-7 integrin - mouse  
C:Species: *Mus musculus* (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 29-Sep-1999  
C:Accession: I61186  
R:Zlobert, B.L.; Vu, M.P.; Waleh, N.; Crawford, J.; Lin, C.S.; Kramer, R.H.  
J. Biol. Chem. 268, 26773-26783, 1993  
A:Title: Alternative extracellular and cytoplasmic domains of the integrin alpha 7 su  
A:Reference number: A49691; MUID:94075378  
A:Accession: I61186  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1135 <RES>  
A:Cross-references: GB:L23423; NID:9431415; PIDN:AAA16600.1; PID:9431416  
C:Superfamily: Integrin alpha-2b chain

Query Match 7.2% Score 8; DB 2; Length 1135;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLALV 14  
|||||||  
Db 1047 LTVLALV 1054

RESULT 15  
JC5950  
Integrin alpha-7 chain precursor - human  
C:Species: *Homo sapiens* (man)  
C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 31-Dec-2000  
C:Accession: JC5950; JC5951; A56839; S40147  
R:Leung, E.; Lim, S.P.; Berg, R.; Yang, Y.; Ni, J.; Wang, S.; Kristensen, G.W.  
Biochem. Biophys. Res. Commun. 243, 317-325, 1998  
A:Title: A novel extracellular domain variant of the human integrin alpha 7 subunit g  
A:Reference number: JC5950; MUID:98139911  
A:Accession: JC5950  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1137 <LEU1>  
A:Cross-references: GB:AF032108; NID:92897115; PIDN:AMC39708.1; PID:92897116  
A:Accession: JC5951  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-657, 733-1137 <LEU2>  
A:Cross-references: GB:AF032108  
R:Wang, W.; Wu, W.; Desai, T.; Ward, D.C.; Kaufman, S.J.  
Genomics 26, 563-570, 1995  
A:Title: Localization of the alpha7 integrin gene (ITGA7) on human chromosome 12q13:  
A:Reference number: A56839  
A:Accession: A56839  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 664-666, 668-688, 690-825, 'X', 827-839, 'D', 841-946, 'H', 948, 'Y', 951, 'P', 953-1  
A:Cross-references: EMBL:X74295; NID:9437781

R;Song, W.K.; Wang, W.; Sato, H.; Bielser, D.; Kaufman, S.  
 A:Submitted to the EMBL Data Library, July 1993  
 A:Description: Expression of alpha 7 integrin cytoplasmic domains during skeletal muscle  
 A:Accession: S40147  
 A:Reference number: S40147  
 A:Accession: S40147  
 A:Molecule type: mRNA  
 A:Residues: 1061-1137 <SON>  
 A:Cross-references: EMBL:X74295; NID:g437781; PIDN:CAA52348.1; PID:g437782  
 C:Genetics:  
 A:Gene: GDB:ITGA7  
 A:Cross-references: GDB:131465; OMIM:600536  
 A:Map position: 12q13-12q13  
 C:Superfamily: Integrin alpha-2b chain  
 C:Keywords: glycoprotein; skeletal muscle; transmembrane protein  
 F:1-33/Domain: Signal sequence #status predicted <Sig>  
 F:34-1137/Product: Integrin alpha-7 chain, long splice form #status predicted <MATL>  
 F:34-657/33-1137/Product: Integrin alpha-7 chain, short splice form #status predicted  
 F:1038-1056/Domain: transmembrane #status predicted <TRM>  
 F:86;742,945,981,1001/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.2%; Score 8; DB 2; Length 1137;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVALLV 14  
 |||||  
 Db 1049 LTVALLV 1056

RESULT 16  
 B53283  
 A:Major cat allergen Fel d I beta chain - cat (fragment)  
 C:Species: Felis silvestris catus (domestic cat)  
 C:Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 12-May-1994  
 C:Accession: B53283  
 R:Dufort, O.A.; Carrelra, J.; Nitli, G.; Polo, F.; Lombardero, M.  
 Mol. Immunol. 28, 301-309, 1991  
 A:Title: Studies on the biochemical structure of the major cat allergen Felis domesticus  
 A:Reference number: A53283; MUID:91287714  
 A:Accession: B53283  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-20 <DUF>

Query Match 6.3%; Score 7; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 PIFDYVF 33  
 |||||  
 Db 8 PIFDYVF 14

RESULT 17  
 T20911  
 A:hypothetical protein F14F8.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T20911  
 R:Lloyd, C.  
 submitted to the EMBL Data Library, March 1997  
 A:Reference number: Z19346  
 A:Accession: T20911  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-73 <MIT>  
 A:Cross-references: EMBL:Z292787; PIDN:CA07183.1; GSPDB:GN00023; CESP:F14F8.8  
 A:Experimental source: Clone F14F8  
 C:Genetics:  
 A:Gene: CESP:F14F8.8

A:Map position: 5  
 A:Introns: 16/3; 46/3

Query Match 6.3%; Score 7; DB 2; Length 73;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVALL 13  
 |||||  
 Db 5 LTVALL 11

RESULT 18  
 D83680  
 A:hypothetical protein BH0244 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: D83680  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; I  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: D83680  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-77 <STO>  
 A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03963.1; GSPDB:C  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH0244

Query Match 6.3%; Score 7; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 ELLIDLS 47  
 |||||  
 Db 24 ELLIDLS 30

RESULT 19  
 G46449  
 A:hypothetical protein Trbo [imported] - plasmid RK2  
 C:Species: plasmid RK2  
 C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Jun-2000  
 C:Accession: G46449  
 R:Lessl, M.; Balzer, D.; Pansegrau, W.; Lanka, E.  
 J. Biol. Chem. 267, 20471-20480, 1992  
 A:Title: Sequence similarities between the RP4 Tir2 and the T1 VirB region strongly s  
 A:Reference number: A44020; MUID:93015931  
 A:Accession: G46449  
 A:Contents: incompatibility group P  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-87 <LES>  
 A:Cross-references: GB:M93696; NID:g152554; PIDN:AAA26441.1; PID:g152569  
 A:Note: sequence extracted from NCBI backbone (NCBIP:115789)  
 C:Genetics:  
 A:Genome: plasmid

Query Match 6.3%; Score 7; DB 2; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11  
 |||||  
 Db 62 GALLVLA 68

RESULT 20

675257  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: G75257  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250, MUID:20036896.  
A:Accession: G75257  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-103 <WHI>  
A:Cross-references: GB:AE002086;  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2573  
A:Map position: 1

Query Match 6.3%; Score 7; DB 2; Length 103;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVLA 12  
|||||  
DB 15 ALLVLA 21

RESULT 21  
AE0751  
Probable lipoprotein STY2172 [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium  
C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AE0751  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium  
A:Reference number: AB0502; PMID:11677608.  
A:Accession: AE0751  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-137 <PAR>  
A:Cross-references: GB:AL51382; PIDN:CAD05712.1; PID:gl6503205; GSPDB:GN0176  
C:Genetics:  
A:Gene: STY2172

Query Match 6.3%; Score 7; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11  
|||||  
DB 8 GALLVLA 14

RESULT 22  
G59094  
hypothetical protein PX01-31 - Bacillus anthracis virulence plasmid pX01  
C:Species: Bacillus anthracis  
C>Date: 12-Nov-1999 #sequence\_revision 12-Nov-1999 #text\_change 11-May-2000  
C:Accession: G59094  
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler, J.; Bacteriol. 181, 6509-6515, 1999  
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbored by B. anthracis strain Ames.  
A:Reference number: A59091, MUID:99445483

A:Accession: G59094  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <OKI>  
A:Cross-references: GB:AE065404; NID:94894216; PIDN:AMD32335.1; PID:94894247  
A:Experimental source: strain Sterne  
C:Genetics:  
A:Gene: PX01-31  
A:Genome: plasmid  
C:Superfamily: Bacillus anthracis virulence plasmid pX01 hypothetical protein PX01-31

Query Match 6.3%; Score 7; DB 2; Length 198;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 NEILLDL 46  
|||||  
DB 150 NEILLDL 156

RESULT 23  
B75497  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C:Accession: B75497  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: B75497  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <WHI>  
A:Cross-references: GB:AE001919; GB:AE000513; NID:96458307; PIDN:AAF10195.1; PID:9645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0608  
A:Map position: 1  
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0608

Query Match 6.3%; Score 7; DB 2; Length 203;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11  
|||||  
DB 57 GALLVLA 63

RESULT 24  
D75505  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000  
C:Accession: D75505  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: D75505  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-244 <WHI>  
A:Cross-references: GB:AE001913; GB:AE000513; NID:96458240; PIDN:AAF10136.1; PID:9645  
A:Experimental source: strain R1  
C:Genetics:

A:Gene: DR0554  
A:Map position: 1  
C:Superfamily: Deinococcus radiodurans hypothetical protein DR0554

Query Match 6.3%; Score 7; DB 2; Length 244;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 LVTVAL 13  
| | | | | | |  
DB 54 LVTVAL 60

RESULT 25  
A97381  
hypothetical transcription regulator in BHD 3' region [imported] - Agrobacterium tumefaciens  
C:Species: Agrobacterium tumefaciens  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: A97381  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Onuollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; PMID:11743194  
A:Accession: A97381  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-245 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AMK86002.1; PID:G15155067; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_306  
A:Map position: circular chromosome

Query Match 6.3%; Score 7; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LVTVAL 18  
| | | | | | |  
DB 212 LVTVAL 218

RESULT 26  
AH2598  
transcription regulator, GntR family Atu0182 [imported] - Agrobacterium tumefaciens (str. C)  
C:Species: Agrobacterium tumefaciens  
C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH2598  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, D.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH2598  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-245 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AA141206.1; PID:G17738508; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu0182  
A:Map position: circular chromosome

Query Match 6.3%; Score 7; DB 2; Length 245;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LVTVAL 18  
| | | | | | |  
DB 212 LVTVAL 218

RESULT 27  
T50677  
bacteriorhodopsin [similarity] - "Haloterrigena" sp. (strain arg-4)  
C:Species: "Haloterrigena" sp.  
A:Variety: strain arg-4  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 03-Nov-2000  
C:Accession: T50677  
R:Ihara, K.; Umemura, T.; Katagiri, I.; Kitajima-Ihara, T.; Sugiyama, Y.; Kimura, Y.; J. Mol. Biol. 285, 163-174, 1999  
A:Title: Evolution of the archaeal rhodopsins: Evolution rate changes by gene duplication  
A:Reference number: Z2703; PMID:99096913  
A:Accession: T50677  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-250 <IHA>  
A:Cross-references: EMBL:AB09620; PIDN:BAW75200.1  
C:Genetics:  
A:Gene: bop  
C:Superfamily: bacteriorhodopsin  
C:Keywords: chromoprotein; photoreceptor; retinal; transmembrane protein  
F:225/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 6.3%; Score 7; DB 2; Length 250;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LITDLS 48  
| | | | | | |  
DB 101 LITDLS 107

RESULT 28  
T46454  
hypothetical protein DKFZp434J212.1 - human  
C:Species: Homo sapiens (man)  
C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46454  
R:Ansorge, W.; Winkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23038  
A:Accession: T46454  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-253 <AAA>  
A:Cross-references: EMBL:AL137395  
A:Experimental source: adult testis; clone DKFZp434J212  
C:Genetics:  
A:Note: DKFZp434J212.1  
C:Superfamily: human hypothetical protein DKFZp434J212.1

Query Match 6.3%; Score 7; DB 2; Length 253;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LVTVAL 13  
| | | | | | |  
DB 167 LVTVAL 173

RESULT 29  
AH1601  
phosphatidate cytidyltransferase (CDP-diglyceride synthase) homolog cdsA [imported]  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AH1601  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baguero, F.; Berche, P.; Bloec

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.  
D.; Jones, L.M.; Karsberg, O.  
Science 294, 843-852, 2001  
A:Authors: Kref, U.; Kohn, M.; Kunst, F.; Kurupkat, G.; Madueno, E.; Maltournam, A.; Me  
ok, C.; Schlueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,  
A:Title: Comparative genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:2137279; PMID:11679669  
A:Accession: AB1601  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CA956584.1; PID:916413826; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: cdsA  
C:Superfamily: phosphatidate cytidyllyltransferase

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 262;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVLLALV 14  
DB 201 LVLLALV 207

RESULT 30  
T36287  
Probable integral membrane protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000  
C:Accession: T36287  
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: 221603  
A:Accession: T36287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-266 <SEB>  
A:Cross-references: EMBL:AL049819; PIDN:CA842664.1; GSPDB:GN00070; SCOEDB:SC07.05C  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC07.05C  
C:Superfamily: Streptomyces penicillin daunorubicin resistance protein

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 266;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALVLLAL 12  
DB 250 ALVLLAL 256

RESULT 31  
AB4746  
Probable NAM (no apical meristem)-like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: AB4746  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487  
A:Accession: AB4746  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-268 <STO>  
A:Cross-references: GB:AE002093; NID:92459430; PIDN:AA80665.1; GSPDB:GN00139

C:Genetics:  
A:Gene: Atg33480  
A:Map position: 2

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 268;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 TCPIFYD 31  
DB 197 TCPIFYD 203

RESULT 32  
S75087  
Hypothetical protein slr0267 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S75087  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O.; K.; Okumura, S.; Shimpo, S.; Shinozaki, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*  
s.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75087  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-279 <KAN>  
A:Cross-references: EMBL:D90910; GB:AB001339; NID:91652956; PIDN:BA17949.1; PID:9165  
C:Superfamily: conserved hypothetical protein MG442

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 279;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 LLLDLST 48  
DB 213 LLLDLST 219

RESULT 33  
T00479  
Probable phosphatidylinositol-glycan synthase [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F1913.21  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001  
C:Accession: T00479; B84763  
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
submitted to the EMBL Data Library, April 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.  
A:Reference number: 214160  
A:Accession: T00479  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-303 <ROU>  
A:Cross-references: EMBL:AC004238; NID:93033373; PID:93033393  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487  
A:Accession: B84763  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-303 <STO>  
A:Cross-references: GB:AE002093; NID:93033393; PIDN:AA12837.1; GSPDB:GN00139

C:Genetics:  
A:Gene: AT2g34980: F1913.21  
A:Map position: 2

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 303;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LLLDLSL 48  
| | | | | | | |  
DB 89 LLLDLSL 95

## RESULT 34

bo-type ubiquinol oxidase (EC 1.10.3.-) chain II goxa - Bacillus subtilis  
N:Alternate names: quinol oxidase aq3-600 chain goxa  
C:Species: Bacillus subtilis

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: E69687; A38129; S39692  
R:Kunster, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berten  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabbri, C.; Ferrar, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallier  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadala, Y.; Sato, T.; Scanlon,  
A.; Authors: Schleich, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MID:98044033  
A:Accession: E69687

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-321 <RUN>  
A:Cross-references: GB:299123; GB:AL009126; MID:g2636240; PIDN:CAB15843.1; PID:g2636352  
A:Experimental source: strain 168  
R:Sentana, M.; Kunst, F.; Hullo, M.F.; Rapoport, G.; Danchin, A.; Glaeser, P.  
J. Biol. Chem. 267, 10225-10231, 1992

A:title: Molecular cloning, sequencing, and physiological characterization of the gox  
A:Reference number: A38129; MID:92268053  
A:Accession: A38129

A:Molecule type: DNA  
A:Residues: 'MHRFRITSGRMV', 2-38, 'D', 39-321 <SNA>

A:Cross-references: GB:M86548; MID:g143395; PIDN:AAA22686.1; PID:g143396  
A:Note: sequence extracted from NCBI backbone (NCBI:103632, NCBI:103599)  
R:Glaeser, P.; Kunst, F.; Annaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,  
A.; Rapoport, G.; Danchin, A.  
Mol. Microbiol. 10, 371-384, 1993

A:title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr  
A:Reference number: S39655; MID:95020537  
A:Accession: S39692

A:Molecule type: DNA  
A:Residues: 'V', 2-38, 'D', 39-321 <GLA>

A:Cross-references: EMBL:X73124  
C:Genetics:

A:Gene: goxa  
C:Superfamily: bo-type ubiquinol oxidase chain II precursor; cytochrome-c oxidase chain  
C:Keywords: copper; electron transfer; membrane-associated complex; oxidoreductase; resp  
F;31-229/Domain: cytochrome-c oxidase chain II homology <CO2>

Query Match  
Best Local Similarity 100.0%; Score 7; DB 1; Length 321;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LTVLALL 13

DB 11 LTVLALL 17  
| | | | | | | |

## RESULT 35

B83579  
hypothetical protein PA0543 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83579

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; I  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MID:20437337

A:Accession: B83579

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <STO>

A:Cross-references: GB:AE004490; GB:AE004091; MID:g9946398; PIDN:AA03932.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:  
A:Gene: PA0543

Query Match  
Best Local Similarity 100.0%; Score 7; DB 2; Length 322;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LTVLALL 13  
| | | | | | | |  
DB 14 LTVLALL 20

## RESULT 36

G64125  
hemlin transport protein homolog H11471 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000  
C:Accession: G64125

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, J.M.; Weidman  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M  
Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MID:95350630

A:Accession: G64125

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-337 <TIGR>

A:Cross-references: GB:U32825; GB:I42023; MID:g3212226; PIDN:AA023119.1; PID:g1574312

C:Superfamily: ferriochrome ABC transporter

Query Match  
Best Local Similarity 100.0%; Score 7; DB 1; Length 337;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALGVKKA 23  
| | | | | | | |  
DB 230 ALGVKKA 236

## RESULT 37

C72403  
hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: C72403

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316  
A:Accession: C72403  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <ARN>  
A:Cross-references: GB:AE001707; GB:AE000512; NID:g4980720; PIDN:AA035332.1; PID:g498073  
A:Experimental source: strain M898  
C:Genetics:  
A:Gene: TM0241

Query Match 6.3%; Score 7; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALLV 14  
DB 322 LVALLV 328

RESULT 38  
S75065  
sensoy transduction histidine kinase sll1590 - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein sll1590  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S75065  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S75065  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-350 <KAN>  
A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BA017927.1; PID:d101866  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG

Query Match 6.3%; Score 7; DB 2; Length 350;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VLLALV 15  
DB 86 VLLALV 92

RESULT 39  
JC1249  
peroxidase (EC 1.11.1.7) BP-2A precursor - barley  
C:Species: *Hordeum vulgare* (barley)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 11-Jun-1999  
C:Accession: JC1249  
R:Tréllade, B.; Rasmussen, S.K.  
Gene 118, 261-266, 1992  
A:Title: Structure and chromosomal localization of the gene encoding barley seed peroxid  
A:Reference number: JC1249; MUID:92380512  
A:Accession: JC1249  
A:Molecule type: DNA  
A:Residues: 1-355 <THE>  
A:Cross-references: GB:M83671; NID:g167082; PIDN:AAA2974.1; PID:g167083  
A:Experimental source: seed  
C:Genetics:

A:Gene: Prx6  
A:Map position: 3  
A:Introns: 85/3  
C:Superfamily: peroxidase  
C:Keywords: oxidoreductase  
F:1-36//Domain: signal sequence #status predicted <Sig>  
F:37-355//Product: peroxidase BP-2A #status predicted <Mat>

Query Match 6.3%; Score 7; DB 2; Length 355;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVAL 12  
DB 13 ALLVAL 19

RESULT 40  
E95126  
amylotransferase, class-V [imported] - *Streptococcus pneumoniae* (strain TIGR4)  
C:Species: *Streptococcus pneumoniae*  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: E95126  
R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
son, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: E95126  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-360 <KOR>  
A:Cross-references: GB:AE005672; PIDN:AAK75206.1; PID:g14972569; GSPDB:GN00164; TIGR:  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP1094  
C:Superfamily: nitrogen fixation protein nifH

Query Match 6.3%; Score 7; DB 2; Length 360;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 QNTVEDL 104  
DB 341 QNTVEDL 347

RESULT 41  
S14355  
peroxidase (EC 1.11.1.7) BP-2B - barley  
C:Species: *Hordeum vulgare* (barley)  
C:Date: 03-Mar-1994 #sequence\_revision 10-Nov-1995 #text\_change 11-Jun-1999  
C:Accession: S14355  
R:Tréllade, B.; Rasmussen, S.K.  
submitted to the EMBL Data Library, June 1993  
A:Description: Barley seed peroxidase BP 2 allele.  
A:Reference number: S14355  
A:Accession: S14355  
A:Molecule type: DNA  
A:Residues: 1-364 <THE>  
A:Cross-references: EMBL:Z23131; NID:g312691; PIDN:CAA80667.1; PID:g312692  
C:Genetics:  
A:Gene: prx6  
A:Introns: 85/3  
C:Superfamily: peroxidase  
C:Keywords: oxidoreductase

Query Match 6.3%; Score 7; DB 2; Length 364;



Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALVIAL 12  
| | | | | | |  
Db 13 ALVIAL 19

## RESULT 42

A:97997

pyridoxal-phosphate dependent aminotransferase [imported] - Streptococcus pneumoniae (st  
C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97997

R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.; H  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97997

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-371 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99805.1; PID:g15458617; GSPDB:GN00174

C:Genetics:

A:Gene: nifs

C:Superfamily: nitrogen fixation protein nifs

Query Match 6.3%; Score 7; DB 2; Length 371;  
Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 QNTVEDL 104  
| | | | | | |  
Db 352 QNTVEDL 358

## RESULT 43

B82488

hypothetical protein VCA0200 [imported] - Vibrio cholerae (strain N16961 serogroup O1).  
C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: B82488

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833.

A:Accession: B82488

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-374 <HEI>

A:Cross-references: GB:AE004360; GB:AE003853; NID:g9657590; PIDN:AAF96613.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0200

A:Map position: 2

Query Match 6.3%; Score 7; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LVALLVT 15  
| | | | | | |  
Db 254 LVALLVT 260

## RESULT 44

E83506  
Probable MFS transporter Pali08 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: E83506

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I  
.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p  
A:Reference number: A82950; MUID:20437337

A:Accession: E83506

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1385 <STO>

A:Cross-references: GB:AE004541; GB:AE004091; NID:g9947024; PIDN:AA04497.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: Pali08

Query Match 6.3%; Score 7; DB 2; Length 385;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LVIALL 13  
| | | | | | |  
Db 10 LVIALL 16

## RESULT 45

H70904

Probable lprK protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70904

R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,  
Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: H70904

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-390 <COL>

A:Cross-references: GB:Z97050; GB:AL123456; NID:g3256008; PIDN:CA09756.1; PID:g22135

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: lprK

Query Match 6.3%; Score 7; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LVIALL 13  
| | | | | | |  
Db 21 LVIALL 27

## RESULT 46

S09813

hypothetical protein U50 - human cytomegalovirus (strain AD169)

N:Alternate names: hypothetical protein HRF4

C:Species: human cytomegalovirus, human herpesvirus 5

A:Variety: strain AD169

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999

C:Accession: S09813

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.; Cerny, R.; Horsnell, T  
M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A:Reference number: S07479; M01D:90269039  
A:Accession: S09813  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-397 <CHE>  
A:Cross-references: EMBL:X17403; NID:q59591; PDB:CA35409.1; PID:g1780828  
A:Experimental source: strain AD169; host Homo sapiens  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1989  
A:Keywords: glycoprotein; transmembrane protein  
A:362-382/Domain: transmembrane [status predicted] <TM>  
A:39/109/Binding site: carbohydrate (asn) (covalent) [status predicted]

Query Match	6.38;	Score 7;	DB 2;	Length 397;	
Best Local Similarity	100.0%;	Pred. No. 59;			
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	7	LLVALL	13		
Db	374	LLVALL	380		

RESULT 47  
 B70621  
 Probable aryl protein - Mycobacterium tuberculosis (strain H37Rv)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #text:revision 17-Jul-1998 #text:change 20-Jun-2000  
 C:Accession: B70621  
 R:Colo: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rastandean, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393: 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.F.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; PMID:9829587  
 A:Accession: B70621  
 A:Status: preliminary: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-400 <COL>  
 A:Cross-references: GB:865982; GB:A122445; NID:g3261718; PIDN:CAB06649.1; PID:g1839007  
 A:Experimental source: strain H37Rv  
 A:Genetics:  
 A:Gene: aryl  
 A:Superfamily: ornithine-oxo-acid aminotransferase

Query Match	Score 7;	DB 2;	Length 400;
Best Local Similarity	6.3%;	Pred. No. 59;	
Matches 7; Conservative	100.0%;	Mismatches 0;	Indels 0;
QY 4 RGALLVL 10			
db 217 RGALLVL 223			

<p>RESULT 48</p> <p>m35334</p> <p>probable membrane protein - Streptomyces coelicolor</p> <p>C:Species: Streptomyces coelicolor</p> <p>C:Date: 03-Nov-1999 #sequence_revision 05-Nov-1999 #date_change 05-Nov-1999</p> <p>C:Accession: J35334</p> <p>R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, May 1999</p> <p>A:Reference number: Z21575</p> <p>A:Accession: J35334</p> <p>A:Status: preliminary; translated from GB/EMBL/DDAJ</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-400 &lt;Olig&gt;</p> <p>A:Cross-references: EMBL:AL049863; PDB:CAE42943.1; GSPDB:GN00070; SCODEB:SC5H1.18</p> <p>A:Experimental source: strain A3(2)</p> <p>C:Genetics:</p> <p>C:Gene: SCODEB:SC5H1.18</p>	
--	--

Query Match	6.38;	Score 7;	DB 2;	Length 400;
Best Local Similarity	100.0%;	Pred. No. 59;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	8	LVALLV	14	
DB	164	LVALLV	170	

RESULT 49  
 G83503  
 C:Probable MFS transporter Pali13 [!imported] - Pseudomonas aeruginosa (strain PAO1)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #extLchange 31-Dec-2000  
 C:Accession: G83503  
 R:Stover, C.K.; Pham, X.O.; Ertvin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 adaman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Kas, A.; Lardig, K.; L  
 .; Lory, S.; Olson, M.V  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: G83503  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1422 <STO>  
 A:Cross-references: GB:AE004543; GB:AE004091; MID:g9947047; PIDN:AMG04520.1; GSPDB:GN  
 A:Experimental source: strain PAO1  
 C:Genetics:  
 C:Gene: Pali13

```

Query Match Similarity      5.38; Score 7; DB 2; Length 422;
Best Local Similarity      100.08; Fred. No. 62;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
QY      5 GALTLYLA 11
      |||||
Db      397 GALTLYLA 403

RESULT 50
S32488
dihydrolipoamide S-acyltransferase (EC 2.3.1.-), alpha-oxo acid dehydrogenase complex;
N:Alternate names: branched-chain alpha-oxo acid dehydrogenase complex B2 component;
C:Species: Bacillus subtilis
C>Date: 02-Dec-1993 #sequence_revision 20-Feb-1995 #ext_change 20-Jun-2000
C:Accession: S32488; E69593
R:Wang, G.F.; Kuriki, T.; Roy, K.L.; Kaneda, T.
Eur. J. Biochem. 213, 1091-1099, 1993
A:title: The primary structure of branched-chain alpha-oxo acid dehydrogenase from Bacillus
Reference number: S32486; MUID: 93279308

```

RESULT 50  
 532488  
 dithiolipomamide S-acyltransferase (EC 2.3.1.-), alpha-oxo acid dehydrogenase complex  
 N-Alkylmalonate names: branched-chain alpha-oxo acid dehydrogenase complex E2 component,  
 C.Species Bacillus pasteurii  
 C.Date: 02-Dec-1995  
 C.Accession: S32488 Sequence-Revision 20-Feb-1995 #text\_change 20-Jun-2000  
 R.Wang, G.F., Kurita, T., Roy, K.L., Kaneeda, T.  
 Eur J. Biochem. 213, 1091-1099, 1993  
 Article: The primary structure of branched-chain alpha-oxo acid dehydrogenase from *Ba*  
 Reference number: S32486; M0ID: J93279308  
 Accession: S32488  
 Molecule type: DNA  
 M.Residue(s): 1-424 <MAN>  
 A.Cross-references: GI:M97391; GI:M96937; NID:9142610; PIDN:AAA22280.1; PID:9142613  
 R.Experimental source: Strain 1685  
 R.Kunitz, F.; Ogasawara, N.; Moszser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 A.; Brion, S.; Boulliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A.Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M  
 Koether, P.; Koningsberger, G.; Krogh, S.; Kunano, M.; Kurita, K.; Lapdus, A.; Lardino  
 A.; Muthers, L.; Pauler, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudgaa, A.; Oudgaa, B.; Park, S.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivolta, C.; Rocha, A.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani  
 A.; Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terstapa, P.; Tognoni, A.; Tosato, Y.; Uchida  
 T.; Winters, P.; Wipfl, A.; Yamano, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A.; Yoshida, Y.; Yoshikawa, H.F.; Zumbst, E.; Yoshikawa, H.; Danchin, A.  
 Article: The complete genome sequence of the Gram-positive bacterium *Bacillus subtili*

A:Reference number: A69580; MUID:98044033  
A:Accession: E69593  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-424 <RUN>  
A:Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CAB14334.1; PID:g2634837  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: bimbB  
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology  
C:Keywords: acyltransferase; fatty acid biosynthesis; oxidoreductase  
F:5-78/Domain: lipoyl/biotin-binding homology <LBP>  
F:395,399/Active site: His, Asp #status predicted

Query Match 6.3%; Score 7; DB 2; Length 424;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 RVLDEL 82  
|||||||  
DB 396 RVLDEL 402

RESULT 51  
E83482  
probable MFS transporter PA1313 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83482  
R:Stover, C.K.; Pham, X.O.; Elwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Linn, N.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83482  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-452 <STO>  
A:Cross-references: GB:AE004560; GB:AE004091; NID:g9947240; PIDN:ANG04702.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1313

Query Match 6.3%; Score 7; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
|||||||  
DB 170 ALLVIAL 176

RESULT 52  
B86722  
biotin carboxylase (EC 6.3.4.14) [imported] - Lactococcus lactis subsp. lactis (strain I  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: B86722  
R:Pollock, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21255186; PMID:11337471  
A:Accession: B86722  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <STO>  
A:Cross-references: GB:AE005176; PID:g12723697; PIDN:AAK04876.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: accC

C:Superfamily: biotin carboxylase; biotin carboxylase homology  
C:Keywords: ligase

Query Match 6.3%; Score 7; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 VANGNEL 42  
|||||||  
DB 316 VANGNEL 322

RESULT 53  
F82412  
probable aspartokinase VCA0822 [imported] - Vibrio cholerae (strain N16961 serogroup  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82412  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: F82412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <HEI>  
A:Cross-references: GB:AE004410; GB:AE003853; NID:g9658244; PIDN:AAF96720.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VCA0822  
A:Map position: 2

Query Match 6.3%; Score 7; DB 2; Length 474;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 LKINTLG 110  
|||||||  
DB 152 LKINTLG 158

RESULT 54  
T34504  
hypothetical protein ZK1290.12 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34504  
R:Raich, A.  
submitted to the EMBL Data Library, July 1995  
A:Description: The sequence of C. elegans cosmid ZK1290.  
A:Reference number: 221535  
A:Accession: T34504  
A:Status: preliminary; translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-484 <TAI>  
A:Cross-references: EMBL:021308; PIDN:AA93321.1; GSPDB:GN00020; CESP:ZK1290.12  
A:Experimental source: strain Bristol N2; clone ZK1290  
C:Genetics:  
A:Gene: CESP:ZK1290.12  
A:Map position: 2  
A:Introns: 66/3; 124/2; 181/2; 392/1

Query Match 6.3%; Score 7; DB 2; Length 484;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 LVMTTIS 87  
|||||||

Db 393 LWMITIS 399

## RESULT 55

T24201  
hypothetical protein R12G8.2 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24201

R:Rasmussen

submitted to the EMBL Data Library, April 1997

A:Reference number: Z19853

A:Accession: T24201

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-485 &lt;MID&gt;

A:Cross-references: EMBL:Z93782; PDB:CA07854.1; GSPDB:GN00023; CESP:R12G8.2

A:Experimental source: clone R12G8

C:Genetics:

A:Gene: CESP:R12G8.2

A:Map position: 5

A:Introns: 74/1; 162/3; 210/1; 288/1; 339/3; 426/3

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 485;  
Matches 7; Conservativity 0; Pred. No. 70; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12

Db 95 ALLVIAL 101

## RESULT 56

S75340  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - *Synechocystis* sp. (strain PCC 6803)

N:Alternate names: protein slr2009

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S75340

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S75340

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-495 &lt;KAN&gt;

A:Cross-references: EMBL:D90904; GB:AB001339; NID:q1652225; PDB:BA017254.1; PID:q165233

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: ndhF

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein M01309

C:Keywords: NAD; oxidoreductase

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 495;  
Matches 7; Conservativity 0; Pred. No. 71; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12

Db 18 ALLVIAL 24

## RESULT 57

S44982  
flagellin - *Shigella sonnei*C:Species: *Shigella sonnei*

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 20-Jun-2000

C:Accession: S44982; S44981

R:Tomimaga, A.; Mahmoud, M.A.H.; Mukahara, T.; Enomoto, M.

Mol. Microbiol. 12, 277-285, 1994

A:Title: Molecular characterization of intact, but cryptic, flagellin genes in the ge

A:Reference number: S44980; MUID:94335647

A:Accession: S44982

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-524 &lt;TON&gt;

A:Cross-references: EMBL:D16821; NID:q391893; PDB:BA04095.1; PID:q391894; EMBL:D168

C:Superfamily: flagellin

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 524;  
Matches 7; Conservativity 0; Pred. No. 74; Mismatches 0; Indels 0; Gaps 0;

OY 48 LTRVNAF 54

Db 196 LTRVNAF 202

## RESULT 58

C75418  
ribonucleoprotein Ro/SS-A-related protein - *Deinococcus radiodurans* (strain R1)C:Species: *Deinococcus radiodurans*

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: C75418

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896

A:Accession: C75418

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-531 &lt;WHI&gt;

A:Cross-references: GB:AE001973; GB:AE00513; NID:g6459001; PDB:AA010833.1; PID:g645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1262

A:Map position: 1

## Query Match

Best Local Similarity 100.0%; Score 7; DB 2; Length 531;  
Matches 7; Conservativity 0; Pred. No. 75; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12

Db 99 ALLVIAL 105

## RESULT 59

E83032  
probable chemotaxis transducer PA4915 [imported] - *Pseudomonas aeruginosa* (strain PAO)C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: E83032

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: E83032

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-541 &lt;STO&gt;

A:Cross-references: GB:AE004904; GB:AE004091; NID:g9951184; PDB:AA08300.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:



C:Genetics:  
A:Gene: vlecA; VC1652  
A:Map position: 1  
C:Keywords: signal transduction

Query Match 6.3%; Score 7; DB 2; Length 584;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 RCALLVL 10  
Db 470 RCALLVL 476

# RESULT 64

hypothetical protein ECS0081 [Imported] - Escherichia coli (strain O157:H7, substrain R1  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence-revision 18-Jul-2001 #text-change 03-Aug-2001  
C:Accession: A90639  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Sumaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8: 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A:Reference number: A90639; MUID:21156231; PMID:11258796  
A:Accession: A90639; MUID:21156231; PMID:11258796  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-588 <NAV>  
A:Cross-references: GB:BA000077; PIDN:BA033504.1; PID:913359537; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 0509952  
C:Genetics:  
A:Gene: ECS0081  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain H

Query Match 6.3%; Score 7; DB 2; Length 588;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 17 ALGVKMA 23  
Db 444 ALGVKMA 450

# RESULT 65

hypothetical protein ilvI [Imported] - Escherichia coli (strain O157:H7, substrain EDL93  
A:Accession: A85490  
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001  
C:Species: Escherichia coli  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
NATURE 409: 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A85490  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-602 <SMO>  
A:Cross-references: GB:AE005174; NID:912512777; PIDN:AA654381.1; GSPDB:GN00145; UWGP:200  
A:Experimental source: strain O157:H7, substrain EDL93  
C:Genetics:  
A:Gene: ilvI  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain H

Query Match 6.3%; Score 7; DB 2; Length 602;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 17 ALGVKMA 23

Db 458 ALGVKMA 464

# RESULT 66

YCEC31  
acetolactate synthase (EC 4.1.3.18) III large chain - Escherichia coli  
N:Alternate names: acetylhydroxy-acid synthase III large chain  
C:Species: Escherichia coli  
C:Date: 28-Aug-1985 #sequence-revision 05-Dec-1997 #text-change 05-Nov-1999  
C:Accession: E64729; S14385; S40590; A01113; I41305  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.,  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277: 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: E64729  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-604 <BLAT>  
A:Cross-references: GB:AE000118; GB:U00096; NID:91786262; PIDN:ACG73188.1; PID:917862  
A:Experimental source: strain K-12, substrain MG1655  
R:Avaya, J.A.  
submitted to the EMBL Data Library, January 1991  
A:Reference number: S14385  
A:Accession: S14385  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 31-231, 'SV', 234-235, 'V', 237-283, 'V', 285-451, 'S', 453-536, 'RG', 540-552, 'G',  
A:Cross-references: EMBL:X55034  
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Miz  
submitted to the EMBL Data Library, December 1992  
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the  
A:Reference number: S40531  
A:Accession: S40590  
A:Molecule type: DNA  
A:Residues: 31-231, 'SV', 234-235, 'V', 237-283, 'V', 285-451, 'S', 453-536, 'RG', 540-552, 'G',  
A:Cross-references: EMBL:D10483  
R:Squires, C.H.; DePelle, M.; Devereux, J.; Calvo, J.M.  
Nucleic Acids Res. 11, 5299-5313, 1983  
A:Title: Molecular structure of ilvIH and its evolutionary relationship to ilvG in Es  
A:Reference number: A93482; MUID:83272971  
A:Accession: A01113  
A:Molecule type: DNA  
A:Residues: 31-231, 'SV', 234-235, 'V', 237-283, 'V', 285-451, 'S', 453-536, 'RG', 540-552, 'G',  
A:Cross-references: GB:X01609; NID:941794; PIDN:CAA25755.1; PID:941795  
J:Haughn, G.W.; Squires, C.H.; DePelle, M.; Largo, C.T.; Calvo, J.M.  
J. Bacteriol. 163, 186-198, 1985  
A:Title: Unusual organization of the ilvIH promoter of Escherichia coli.  
A:Reference number: I41305; MUID:85234358  
A:Accession: I41305  
A>Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 31-38 <RES>  
A:Cross-references: GB:M10738; NID:9146472; PIDN:AAA24026.1; PID:9146473  
C:Comment: Isozyme III is one of the isozymes sensitive to valine inhibition.  
C:Genetics:  
A:Gene: ilvI  
A:Map position: 2 min  
C:Complex: heterodimer; composed of large chain and small chain  
C:Function:  
A:Description: catalyzes formation of acetylhydroxy acids from the corresponding alpha  
A:Pathway: branched-chain amino acid biosynthesis  
A>Note: thiamine pyrophosphate cofactor; magnesium required  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domai  
C:Keywords: branched-chain amino acid biosynthesis; carbon-carbon lyase; flavoprotein  
F:467-515/domain: thiamin pyrophosphate-binding domain homology <FPD>  
F:81/Active site: Glu #status predicted

Query Match 6.3%; Score 7; DB 1; Length 604;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ALGVKMA 23  
|||||||  
Db 460 ALGVKMA 466

RESULT 67  
F83779  
Iron (III) dicitrate transport system (permease) BH1038 [imported] - Bacillus halodurans  
C:Species: Bacillus halodurans  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83779  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MIMD:20512582; PMID:11058132  
A:Accession: F83779  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-649 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04757.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH1038

Query Match 6.3%; Score 7; DB 2; Length 649;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVAL 12  
|||||||  
Db 120 ALLVAL 126

RESULT 68  
F83228  
Hypothetical protein PA3340 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83228  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mlczoguchl, S.D.; Warrenner, P.; Hickey, M.J.; Bt  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Llm,  
.; Lory, S.; Olson, M.V  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MIMD:20437337  
A:Accession: F83228  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-682 <STO>  
A:Cross-references: GB:AE004756; GB:AE004091; NID:g9949466; PIDN:AAG06728.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA3340

Query Match 6.3%; Score 7; DB 2; Length 682;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLAL 13  
|||||||  
Db 360 LTVLAL 366

RESULT 69  
A53796  
gelatinase B (EC 3.4.24.35) precursor - rabbit  
N:Alternate names: 92K matrix metalloproteinase; 92K type IV collagenase; matrix metallo  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A53796; A55398

R:Tezuka, K.; Nemoto, K.; Tezuka, Y.; Sato, T.; Ikeda, Y.; Kobori, M.; Kawashima, H.;  
J. Biol. Chem. 269, 15006-15009, 1994  
A:Title: Identification of matrix metalloproteinase 9 in rabbit osteoclasts.  
A:Reference number: A53796; MIMD:94253056  
A:Accession: A53796  
A:Molecule type: mRNA  
A:Residues: 1-707 <TE2>  
A:Cross-references: GB:D26514; NID:g499372; PIDN:BA05520.1; PID:g499373  
A:Experimental source: osteoclasts  
R:Finl, M.E.; Bartlett, J.D.; Matsubara, M.; Rinehart, W.B.; Mody, M.K.; Girard, M.T.  
J. Biol. Chem. 269, 28620-28628, 1994  
A:Title: The rabbit gene for 92-kDa matrix metalloproteinase. Role of AP1 and AP2 in  
A:Reference number: A55398; MIMD:95050662  
A:Accession: A55398  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-75; 'P', 77-99, 'ASR', 103-171 <FIN>  
A:Cross-references: GB:L36050; NID:g535714; PIDN:AAA64358.1; PID:g535715  
C:Genetics:  
A:Introns: 46/3; 124/2  
C:Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat h  
C:Keywords: glycoprotein; hydrolase; metalloproteinase; zinc; zymogen  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-106/Domain: activation peptide #status predicted <PRO>  
F:107-707/Product: 91K neutrophil gelatinase B #status predicted <MAT>  
F:230-271/Domain: fibronectin type II repeat homology <2P9>  
F:288-329/Domain: fibronectin type II repeat homology <2P8>  
F:347-388/Domain: fibronectin type II repeat homology <2P1>  
F:510-704/Domain: hemopexin repeat homology (Asn) (covalent) #status predicted  
F:88,120,127/Binding site: carbohydrate (Asn)  
F:99,401,405,411/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #stat  
F:401,405,411/Binding site: zinc, catalytic (His) (active) #status predicted  
F:402/Active site: Glu #status predicted

Query Match 6.3%; Score 7; DB 1; Length 707;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LTVLAL 14  
|||||||  
Db 7 LTVLAL 13

RESULT 70  
I46031  
gelatinase B (EC 3.4.24.35) - bovine  
N:Alternate names: matrix metalloproteinase 9 (MMP9)  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: I46031; S43112  
R:Baylis, H.A.; Meson, A.; Hall, R.  
Mol. Biochem. Parasitol. 69, 211-222, 1995  
A:Title: Infection with Theileria annulata induces expression of matrix metalloprote  
A:Reference number: I46031; MIMD:95287902  
A:Accession: I46031  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-712 <BAV>  
A:Cross-references: EMBL:X78324; NID:g467620; PIDN:CAA5127.1; PID:g467621  
A:Note: submitted to the EMBL Data Library, March 1994  
C:Genetics:  
A:Gene: mmp9  
C:Superfamily: gelatinase A; fibronectin type II repeat homology; hemopexin repeat ho  
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen  
F:67-216,392-444/Domain: matrix metalloproteinase homology #status atypical <MMP>  
F:230-271/Domain: fibronectin type II repeat homology <2P1>  
F:288-329/Domain: fibronectin type II repeat homology <2P2>  
F:347-388/Domain: fibronectin type II repeat homology <2P3>  
F:515-709/Domain: hemopexin repeat homology <PYN>  
F:99,401,405,411/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #stat  
F:401,405,411/Binding site: zinc, catalytic (His) (active) #status predicted  
F:402/Active site: Glu #status predicted

```

Query Match      6.3%; Score 7; DB 1; Length 712;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVTALV 14
    |||||
Db 7 LVTALV 13

RESULT 71
T23474
hypothetical protein K087.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23474
R:Smyle, R.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19745
A:Accession: T23474
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-729 <MIL>
A:Cross-references: EMBL:Z77666; PIDN:CA801230.1; GSPDB:GN00022; CESP:K087.7
A:Experimental source: clone K087
C:Genetics:
A:Gene: CESP:K087.7
A:Map position: 4
A:Introns: 33/2; 292/2; 405/3; 555/3; 665/3; 686/3

Query Match      6.3%; Score 7; DB 2; Length 729;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 96 AVONTVE 102
    |||||
Db 651 AVONTVE 657

RESULT 72
AE2217
hypothetical protein al13292 [Imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AE2217
R:Kanehko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Itiguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-859 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA874991.1; PID:g17132387; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al13292

Query Match      6.3%; Score 7; DB 2; Length 859;
Best Local Similarity 100.0%; Pred. No. 1,1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GEAYONT 100
    |||||
Db 692 GEAYONT 698

RESULT 73
S12319
pre-mRNA splicing factor PRP6 - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YBR0508; protein YBR055c
C:Species: Saccharomyces cerevisiae
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 21-Jul-2000
C:Accession: S12319; S45913; S49510; S55854
R:Legrain, P.; Choulika, A.
EMBO J. 9, 2775-2781, 1990
A:Title: The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cy
A:Reference number: S12319; MUID:90360988
A:Accession: S12319
A:Molecule type: DNA
A:Residues: 1-899 <LEG>
A:Cross-references: EMBL:X53465; NID:q4238; PIDN:CA937559.1; PID:q4239
R:Aljinovic, G.; Pohl, F.M.; Pohl, T.M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45906
A:Accession: S45913
A:Molecule type: DNA
A:Residues: 1-899 <ALJ>
A:Cross-references: EMBL:Z35924; NID:q536290; PID:g536291; MIPS:YBR055c
A:Experimental source: strain S288C
R:Aljinovic, G.
submitted to the EMBL Data Library, October 1994
A:Description: Sequence and analysis of 24 kb on chromosome II of Saccharomyces cere
A:Reference number: S49503
A:Accession: S49510
A:Molecule type: DNA
A:Residues: 1-899 <ALJ>
A:Cross-references: EMBL:Z46260; NID:g559942; PID:g559951
A:Experimental source: strain S288C
R:Aljinovic, G.; Pohl, T.M.
Yeast 11, 475-479, 1995
A:Title: Sequence and analysis of 24 kb on chromosome II of Saccharomyces cerevisiae.
A:Reference number: S55854; MUID:95321020
A:Accession: S55854
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Residues: 1-899 <ALM>
A:Cross-references: EMBL:Z46260; NID:g559942; PIDN:CA86398.1; PID:g559951
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1994
C:Genetics:
A:Gene: SGD:PRP6
A:Cross-references: SGD:S0000259; MIPS:YBR055c
A:Map position: 2R
C:Keywords: nucleus; RNA binding; zinc finger

Query Match      6.3%; Score 7; DB 2; Length 899;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LVTQAL 18
    |||||
Db 758 LVTQAL 764

RESULT 74
S75577
cation-transporting ATPase (EC 3.6.1.-) pact-3 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr0822
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75577
R:Kanehko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O.K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75577
A:Status: nucleic acid sequence not shown; translation not shown

```



A:Molecule type: DNA  
A:Residues: 1-953 <RAN>  
A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BA18138.1; PID:g165322  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: pacL-3  
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain  
C:Keywords: hydrolase; ion transport  
F:586-756/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 6.3%; Score 7; DB 2; Length 953;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11  
|||||||  
DB 443 GALLVLA 449

RESULT 75  
E84923  
Hypothetical protein At2g48110 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84923  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487  
A:Accession: E84923  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1226 <STO>  
A:Cross-references: GB:AE002093; NID:g6554175; PIDN:AAD13716.2; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g48110  
A:Map position: 2

Query Match 6.3%; Score 7; DB 2; Length 1226;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLALL 13  
|||||||  
DB 84 LTVLALL 90

RESULT 76  
H96630  
protein r3p18.3 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H96630  
R:theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.E.; Hughes, B.; Huzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: H96630  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1309 <STO>  
A:Cross-references: GB:AE005173; NID:g5454189; PIDN:AAD43604.1; GSPDB:GN00141

C:Genetics:  
A:Gene: T3p18.3  
A:Map position: 1  
C:Superfamily: retrovirus-related polyprotein

Query Match 6.3%; Score 7; DB 2; Length 1309;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 TTRSSK 90  
|||||||  
DB 247 TTRSSK 253

RESULT 77  
VCBEH6  
major capsid protein - human herpesvirus 6 (strain Uganda-1102)  
C:Species: human herpesvirus 6  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 06-Jan-1995  
C:Accession: E33560; A33941  
R:Lawrence, G.L.; Chee, M.; Craxton, M.A.; Gompels, U.A.; Honess, R.W.; Barrell, B.G.  
J. Virol. 64, 287-299, 1990  
A:Title: Human herpesvirus 6 is closely related to human cytomegalovirus.  
A:Reference number: A33560; MUID:90080132  
A:Accession: E33560  
A:Molecule type: DNA  
A:Residues: 1-1345 <LAN>  
A:Cross-references: GB:M28243  
R:Blitler, E.; Lawrence, G.; Liu, M.Y.; Barrell, B.G.; Arrand, J.R.  
J. Virol. 64, 714-722, 1990  
A:Title: Identification, cloning, and expression of the major capsid protein gene of  
A:Reference number: A33941; MUID:90112641  
A:Accession: A33941  
A:Molecule type: DNA  
A:Residues: 1-1345 <LIR>  
C:Superfamily: varicella-zoster virus major capsid protein  
C:Keywords: capsid protein; coat protein

Query Match 6.3%; Score 7; DB 1; Length 1345;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DLKINTL 109  
|||||||  
DB 354 DLKINTL 360

RESULT 78  
T44017  
major capsid protein [imported] - human herpesvirus 6 (strain HST)  
C:Species: human herpesvirus 6  
A:Variety: strain HST  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T44017  
R:Isegawa, Y.; Nakai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; K  
J. Virol. 73, 8053-8063, 1999  
A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a  
A:Reference number: 227332; MUID:99412319  
A:Accession: T44017  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1345 <ISB>  
A:Cross-references: EMBL:AB021506; NID:g4995977; PIDN:BA78278.1; PID:g4996045  
A:Experimental source: strain HST; pop. variant B  
C:Genetics:  
A:Note: U57  
C:Superfamily: varicella-zoster virus major capsid protein  
Query Match 6.3%; Score 7; DB 2; Length 1345;



R:Platnev, A.G.; Yamshechikov, V.F.; Blinov, V.M.  
 Virology 174, 250-263, 1990  
 A:Title: Nucleotide sequence of the genome and complete amino acid sequence of the poly  
 A:Reference number: A33776; MUID:90101381  
 A:Accession: A33776  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3412 <PLE>  
 A:Cross-references: GB:X07755  
 R:Platnev, A.G.; Yamshechikov, V.F.; Blinov, V.M.  
 FEBS Lett. 200, 317-321, 1986  
 A:Title: Tick-borne encephalitis virus genome: the nucleotide sequence coding for virion  
 A:Reference number: A91360; MUID:86220766  
 A:Accession: A24055  
 A:Molecule type: genomic RNA  
 A:Residues: 1-62, 'V', 64-101, 'A', 103-508, 'T', 510-683 <PL2>  
 A:Accession: B24055  
 A:Molecule type: genomic RNA  
 A:Residues: 242, 'SG', 245-246, 358-359, 'EHESD', 366, 'A', 758-849, 'D', 851-1002 <PL3>  
 C:Superfamily: Yellow fever virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;  
 F:2-113/Product: capsid protein C #status predicted <CPC>  
 F:113-205/Product: envelope protein prM #status predicted <PRM>  
 F:206-280/Product: envelope protein M #status predicted <PM>  
 F:281-776/Product: major envelope protein E #status predicted <PME>  
 F:777-1190/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1191-1358/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1359-1489/Product: nonstructural protein NS2b #status predicted <NS2b>  
 F:1490-2110/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1671-1678/Region: nucleotide-binding motif A (P-loop)  
 F:1775-1780/Region: nucleotide-binding motif B  
 F:1779-1782/Region: DEAD motif  
 F:2111-2259/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:2260-2510/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2511-3412/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:144,434,641,753,861,983,999,1228,1649,1988,2044,2052,2447,2466,2685,2725/Binding site:  
 F:144,434,641,753,861,983,999,1228,1649,1988,2044,2052,2447,2466,2685,2725/Binding site:

Query Match 6.3%; Score 7; DB 1; Length 3412;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VIALVLT 15  
 |||||  
 Db 1170 VIALVLT 1176

RESULT 84  
 GENE  
 genome polyprotein - tick-borne encephalitis virus (subtype Western, strain Neudoerfl)  
 N:Contains: capsid protein C; envelope protein E; membrane protein M; nonstructural prot  
 In NS4g: nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: tick-borne encephalitis virus  
 C:Date: 31-Dec-1989 #sequence,revision 30-Jun-1991 #text,change 19-Jan-2001  
 C:Accession: A31052; A32596  
 R:Mandl, C.W.; Heinz, F.X.; Kunz, C.  
 Virology 166, 197-205, 1988  
 A:Title: Sequence of the structural proteins of tick-borne encephalitis virus (Western s  
 A:Reference number: A31052; MUID:86322870  
 A:Accession: A31052  
 A:Molecule type: genomic RNA  
 A:Residues: 1-779 <MAN1>  
 A:Cross-references: GB:M21498  
 R:Mandl, C.W.; Heinz, F.X.; Stoekl, E.; Kunz, C.  
 Virology 173, 291-301, 1989  
 A:Title: Genome sequence of tick-borne encephalitis virus (Western subtype) and comparat  
 A:Reference number: A32596; MUID:90051080  
 A:Accession: A32596  
 A:Molecule type: genomic RNA  
 A:Residues: 767-3414 <MAN2>  
 A:Cross-references: GB:M27157  
 C:Superfamily: yellow fever virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;  
 F:2-116/Product: capsid protein C #status predicted <CPC>

F:117-280/Product: membrane protein M precursor #status predicted <MP>  
 F:117-205/Domain: nonterminal signal sequence #status predicted <SIG>  
 F:206-280/Product: membrane protein M #status predicted <MPM>  
 F:246-264/Domain: transmembrane #status predicted <TM1>  
 F:281-776/Product: envelope protein E #status predicted <EPE>  
 F:738-751/Domain: transmembrane #status predicted <TM2>  
 F:777-1128/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1129-1358/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1359-1489/Product: nonstructural protein NS2b #status predicted <NS2b>  
 F:1490-2110/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1688-1695/Region: nucleotide-binding motif A (P-loop)  
 F:1775-1780/Region: nucleotide-binding motif B  
 F:1779-1782/Region: DEAD motif  
 F:2111-2259/Product: nonstructural protein NS4 #status predicted <NS4>  
 F:2260-2510/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2512-3414/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:144,434,641,753,861,983,999,1649,1988,2044,2447,2529,2686,2726/Binding site: carboi  
 F:144,434,641,753,861,983,999,1649,1988,2044,2447,2529,2686,2726/Binding site:

Query Match 6.3%; Score 7; DB 1; Length 3414;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VIALVLT 15  
 |||||  
 Db 1170 VIALVLT 1176

RESULT 85  
 D97553  
 hypothetical protein AGR\_C\_2949 [imported] - Agrobacterium tumefaciens (strain C58, C  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence,revision 30-Sep-2001 #text,change 11-Jan-2002  
 C:Accession: D97553  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
 A.: Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
 A:Reference number: A97559; PMID:11743194  
 A:Accession: D97553  
 A:Molecule type: preliminary  
 A:Residues: 1-35 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK87381.1; PIDN:q15156688; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_2949  
 A:Map position: circular chromosome

Query Match 5.4%; Score 6; DB 2; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALIVIA 11  
 |||||  
 Db 6 ALIVIA 11

RESULT 86  
 AE3461  
 hypothetical protein BME1675 [imported] - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence,revision 01-Feb-2002 #text,change 01-Feb-2002  
 C:Accession: AE3461  
 R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov  
 .; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: AE3461  
 A:Molecule type: preliminary  
 A:Residues: 1-78 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAJ52856.1; PID:g17983698; GSPDB:GN00190  
 A:Experimental source: strain 16M  
 A:Gene: BME11675  
 A:Map position: I

Query Match  
 Best Local Similarity 5.4%; Score 6; DB 2; Length 78;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GRAYON 99  
 DB 53 GRAYON 58

## RESULT 87

Conserved hypothetical protein ML0802 [Imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: D87009  
 R:Coile, S.T.; Eigimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
 Nature 409, 1007-1011, 2001  
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sg  
 A:Title: Massive gene decay in the leprosy bacillus  
 A:Reference number: A86509; MUID:21128732; PMID:11234002

A:Accession: D87009  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-79 <STO>

A:Cross-references: GB:AA50380; NID:g13092898; PIDN:CAC30312.1; GSPDB:GN00147  
 C:Genetics:  
 A:Gene: ML0802

Query Match  
 Best Local Similarity 5.4%; Score 6; DB 2; Length 79;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 GYKMAE 24  
 DB 6 GYKMAE 11

## RESULT 88

Intestinal trefoil factor 3 precursor - human  
 N:Alternate names: HTRF; P-domain peptide hpl.B  
 C:Species: Homo sapiens (man)

C>Date: 02-Jun-1994 #sequence\_revision 02-Jun-1994 #text\_change 16-Jul-1999  
 C:Accession: A48284; A45491; A45979

R:Hauser, F.; Poulsom, R.; Chinery, R.; Rogers, L.A.; Hanby, A.M.; Wright, N.A.; Hoffman  
 Proc. Natl. Acad. Sci. U.S.A. 90, 6961-6965, 1993  
 A:Title: hpl.B, a human P-domain peptide homologous with rat intestinal trefoil factor,  
 A:Reference number: A48284; MUID:93348192

A:Accession: A48284  
 A:Molecule type: mRNA  
 A:Residues: 1-80 <HNU>

A:Cross-references: GB:LI5203; NID:g402482; PIDN:AAA59981.1; PID:g402483  
 R:Podolsky, D.K.; Lynch-Devaney, K.; Stow, J.L.; Oates, P.; Murgue, B.; DeBeaumont, M.;  
 J. Biol. Chem. 268, 6694-6702, 1993

A:Title: Identification of human intestinal trefoil factor. Goblet cell-specific expres  
 A:Reference number: A45491; MUID:93203271

A:Accession: A45491  
 A:Molecule type: mRNA

A:Residues: 8-73, 77-80 <POD>  
 A:Cross-references: GB:LO8044; NID:g307520; PIDN:AAA36766.1; PID:g307521  
 A:Experimental source: colon  
 A:Note: sequence extracted from NCBI backbone (NCBIN:127995, NCBI:127996)  
 C:Genetics:

A:Gene: GDB:TFP3  
 A:Cross-references: GDB:629964; OMIM:600633  
 A:Map position: 21q22.3-21q22.3  
 A:Superfamily: secretory protein xpl; trefoil homology  
 C:Keywords: intestine  
 F.1-21/Domain: signal sequence #status predicted <SIG>  
 F.32-72/Domain: trefoil homology <TRF>

Query Match  
 Best Local Similarity 5.4%; Score 6; DB 2; Length 80;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVLALL 13  
 DB 11 LVLALL 16

## RESULT 89

hypothetical protein all6r - Chlorella virus PCV-1  
 C:Species: Chlorella virus PCV-1  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T17606  
 R:Graves, M.V.; Van Etten, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A:Reference number: 218806

A:Accession: T17606  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-81 <GRA>

A:Cross-references: EMBL:042580; NID:g4028896; PIDN:ACG66484.1  
 A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:  
 A:Note: all6r

Query Match  
 Best Local Similarity 5.4%; Score 6; DB 2; Length 81;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 TISSSK 90  
 DB 48 TISSSK 53

## RESULT 90

hypothetical protein XP1461 [Imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: B82678  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A85515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82678

A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-87 <STM>  
 A:Cross-references: GB:AE003976; GB:AE003849; NID:g9106479; PIDN:AAE84270.1; GSPDB:GN

A:Experimental source: strain 9a5c  
 R:Simpton, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Ungaletta, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; La  
 Chado, M.A.; Madella, H.M.F.; Martino, C.L.; Marques, M.V.; Martins  
 A:Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:genetics:  
A:gene: XFI461

Query Match 5.4%; Score 6; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 2 NELLLD 7

RESULT 91  
A84166  
hypothetical protein Vhg0055h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: A84166  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Tsaky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Hanson, M.J.; Hough, D.W.; Maddocks, D.G.; Jaido Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: A84166  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-88 <STO>  
A:Cross-references: GB:AE004437; NID:g10579707; PIDN:AA618693.1; GSPDB:GN00138  
C:genetics:  
A:gene: VNG0055H

Query Match 5.4%; Score 6; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 75 SRVLDG 80  
45 SRVLDG 50

RESULT 92  
E97828  
hypothetical protein RC1029 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: E97828  
R:Ogata, H.; Audic, S.; Renesco-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: E97828  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-88 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AML03567.1; PID:g15620146; GSPDB:GN00173  
C:genetics:  
A:gene: RC1029

Query Match 5.4%; Score 6; DB 2; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 5 GALLVL 10  
111111

Db 77 GALLVL 82

RESULT 93  
T30720  
hypothetical protein 118L - Molluscum contagiosum virus 1  
N:Alternate names: MC118L  
C:Species: Molluscum contagiosum virus 1  
C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 11-May-2000  
C:Accession: T30720  
R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B. Science 273, 813-816, 1996  
A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host  
A:Reference number: Z20876; MUID:96325459  
A:Accession: T30720  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-94 <SEN>  
A:Cross-references: EMBL:U60315; PIDN:AA55246.1  
C:genetics:  
A:Note: MC118L

Query Match 5.4%; Score 6; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 6 ALLVLA 11  
20 ALLVLA 25

RESULT 94  
T20861  
hypothetical protein F13G3.10 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
R:McMurray, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19334  
A:Accession: T20861  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-96 <WTL>  
A:Cross-references: EMBL:Z71259; PIDN:CAA95797.1; GSPDB:GN00019; CESP:F13G3.10  
A:Experimental source: clone F13G3  
R:Wilkinson, J.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19568  
A:Accession: T22480  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-96 <W12>  
A:Cross-references: EMBL:Z71263; PIDN:CAA95825.1; GSPDB:GN00019; CESP:F13G3.10  
A:Experimental source: clone F52A8  
C:genetics:  
A:Gene: CESP:F13G3.10  
A:Map position: 1  
A:introns: 22/3; 57/3

Query Match 5.4%; Score 6; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 103 DLKINT 108  
111111  
Db 32 DLKINT 37  
RESULT 95  
H88886

Protein F52G2.4 [imported] - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C:Accession: H88886  
 R:anonymus: The C. elegans Sequencing Consortium.  
 A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology  
 A:Reference number: A75000; PMID:9069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_ele  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: H88886  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-98 <STO>  
 A:Cross-references: GB:chr\_IV; PIDD:CAM05208.1; PID:93924777; GSPDB:GN00022; CESP:F52G2.  
 C:Genetics:  
 A:Gene: F52G2.4  
 A:Map position: 4

Query Match 5.4%; Score 6; DB 2; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 DGLVMT 84  
 |||||  
 Db 9 DGLVMT 14

RESULT 96  
 T42908  
 Hypothetical protein Y0075 - *Yersinia pestis* plasmid pCD1  
 C:Species: *Yersinia pestis*  
 C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 21-Jul-2000  
 C:Accession: T42908  
 R:Perry, R.D.; Straley, S.C.; Fetherston, J.D.; Rose, D.J.; Gregor, J.; Blattner, F.R.  
 Infect. Immun. 66, 4611-4623, 1998  
 A:Title: DNA sequencing and analysis of the low-Ca<sup>2+</sup>-response plasmid pCD1 of *Yersinia F*  
 A:Reference number: 222273; PMID:98427122  
 A:Accession: T42908  
 A:Status: preliminary; translated from GB/EMBL/DBD  
 A:Molecule type: DNA  
 A:Residues: 1-99 <PER>  
 A:Cross-references: EMBL:AF074612; NID:93822037; PIDD:AMC09816.1; PID:93822096  
 A:Experimental source: strain KIMS  
 C:Genetics:  
 A:Genome: plasmid pCD1  
 A:Note: Y0075

Query Match 5.4%; Score 6; DB 2; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 LISRYL 78  
 |||||  
 Db 79 LISRYL 84

RESULT 97  
 D86797  
 Prophage p13 protein 06, holin [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1  
 C:Species: *Lactococcus lactis* subsp. *lactis*  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: D86797  
 R:Bolotin, A.; Wincker, P.; Mauget, S.; Jallion, O.; Malarme, K.; Weissenhach, J.; Ehrli  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s  
 A:Reference number: A86625; PMID:21235186; PMID:11337471  
 A:Accession: D86797  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-99 <STO>

A:Cross-references: GB:AE005176; PIDD:q12724366; PIDD:AMC05478.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: p1306

Query Match 5.4%; Score 6; DB 2; Length 99;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVIAL 12  
 |||||  
 Db 14 LTVIAL 19

RESULT 98  
 F70309  
 Protein export membrane protein SecG - *Aquifex aeolicus*  
 C:Species: *Aquifex aeolicus*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: F70309  
 R:Deckerl, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
 A:Reference number: A70300; PMID:9819666  
 A:Accession: F70309  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-100 <AOE>  
 A:Cross-references: GB:AE000673; GB:AE000657; NID:92982834; PIDD:AMC06462.1; PID:9298  
 A:Experimental source: strain VP5  
 C:Genetics:  
 A:Gene: secG  
 A:Superfamily: protein-export protein secG

Query Match 5.4%; Score 6; DB 1; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVIALL 13  
 |||||  
 Db 63 LVIALL 68

RESULT 99  
 E90093  
 Hypothetical protein orf100 [imported] - *Guillardia theta* nucleomorph  
 C:Species: nucleomorph *Guillardia theta*  
 A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
 C:Accession: E90093  
 R:Douglas, S.; Zauner, S.; Franholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;  
 Nature 410, 1091-1096, 2001  
 A:Title: The highly reduced genome of an enslaved algal nucleus.  
 A:Reference number: A99082; PMID:11323671  
 A:Accession: E90093  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-100 <DOU>  
 A:Cross-references: GB:AF165818; NID:q13794502; PIDD:AAK39877.1; GSPDB:GN00150  
 C:Genetics:  
 A:Gene: orf100  
 A:Map position: 1  
 A:Genome: nucleomorph  
 C:Keywords: nucleomorph

Query Match 5.4%; Score 6; DB 2; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LLDLS 47  
 |||||  
 Db 40 LLDLS 45

## RESULT 100

MNVUMB

nonstructural protein NS - Maguari virus

C:Species: Maguari virus

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jun-2000

C:Accession: B33076; J0425

R;Elliott, R.M.; McGregor, A.

Virology 171, 516-524, 1989

A:Title: Nucleotide sequence and expression of the small (S) RNA segment of Maguari buny

A:Reference number: A33076; MUID:89348010

A:Accession: B33076

A:Molecule type: genomic RNA

A:Residues: 1-101 &lt;EL&gt;

A:Cross-references: GB:M28380; GB:D00380; NID:9601857; PIDN:AAA57148.1; PID:9601859

R;Elliott, R.M.

submitted to the EMBL Data Library, January 1989

A:Reference number: J0421

A:Accession: J0425

A:Molecule type: genomic RNA

A:Residues: 1-101 &lt;EL2&gt;

A:Cross-references: GB:D13783; NID:g222091; PIDN:BAA02927.1; PID:g222093

C:Genetics:

A:Map position: segment S

C:Superfamily: bunyavirus nonstructural protein

C:Keywords: nonstructural protein

Query Match 5.4%; Score 6; DB 1; Length 101;  
 Best Local Similarity 100.0%; Pred.No.1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 80 GLVMTT 85  
 |||||  
 Db 27 GLVMTT 32

Search completed: July 9, 2002, 16:18:40  
 Job time: 30 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:19:15 ; Search time 28.88 Seconds

(Without alignments)  
426.911 Million cell updates/sec

Title: US-09-662-784-6

Perfect score: 552

Sequence: 1 DTMRGALVLLALVTQALGV.....CMGEAVQNTVEDKLTFLGR 111

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 135323

Minimum DB seq length: 0

Maximum DB seq length: 10

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

\_A\_Geneseq\_032802:--  
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	8.3	8	21	TRFP chain 2 pept1
2	46	8.3	8	21	Feline human TRFP
3	46	8.3	8	21	Peptide #2. Unide
4	40	7.2	7	21	Feline human TRFP
5	40	7.2	7	21	leader sequence pe
6	40	7.2	7	21	Human TRFP derived
7	40	7.2	7	21	Mutant P53 tumour
8	33	6.0	10	20	TRFP chain 2 pept1
9	32	5.8	6	21	Feline human TRFP
10	32	5.8	6	21	Human TRFP chain 2
11	32	5.8	6	21	Human TRFP chain 2

12	32	5.8	8	22	ACG89573	p53 epitope A2 sup
13	32	5.8	9	17	AA97531	Antigenic peptide, p53 epitope (aa 13
14	32	5.8	9	20	AA53375	Amino acid sequenc
15	32	5.8	9	20	AA40198	HLA-B8 binding p53
16	32	5.8	9	20	AA26714	HLA binding peptid
17	32	5.8	9	21	AA54185	Vaccine related MH
18	32	5.8	9	22	AA49346	Human Leukocyte An
19	32	5.8	9	22	AAU26572	p53 epitope A2 sup
20	32	5.8	9	22	AA89541	p53 epitope A2 sup
21	32	5.8	9	22	AA89572	p53 epitope A2 sup
22	32	5.8	9	22	AA89593	p53 epitope A2 sup
23	32	5.8	9	22	AA89694	p53 epitope HLA-A2
24	32	5.8	10	21	AA54193	p53 epitope HLA-A2
25	32	5.8	10	14	AA44268	Residues 136-145 o
26	32	5.8	10	15	AA61603	Peptide fragment (
27	32	5.8	10	18	AA222030	Antigenic human P5
28	32	5.8	10	21	AA54193	HLA binding peptid
29	32	5.8	10	22	AA49347	Vaccine related MH
30	32	5.8	10	22	AAU26580	Human Leukocyte An
31	32	5.6	9	15	AA85636	Peptide inhibitor
32	32	5.6	9	21	AA80056	N-terminal fragmen
33	32	5.6	9	21	AA54190	HLA binding peptid
34	32	5.6	9	22	AAU26577	Human Leukocyte An
35	32	5.6	9	22	AA89366	p53 DR supermotif
36	32	5.6	9	22	AA89542	p53 epitope A2 sup
37	32	5.6	9	22	AA89594	p53 epitope A2 sup
38	32	5.4	9	21	AA85143	Human secreted pep
39	30	5.4	10	15	AA861597	Peptide fragment (
40	30	5.4	10	20	AA108725	ZIMV N1b protein f
41	29	5.3	10	20	AA108731	PSBMV N1b protein
42	29	5.3	10	22	AA894945	Human complementar
43	29	5.3	10	22	AA84213	Arabidopsis thalia
44	28.5	5.2	9	22	AAU02293	HLA binding TADG-1
45	28	5.1	9	15	AA861691	HLA-A2.1 algorithm

#### ALIGNMENTS

RESULT 1  
AAB28957 standard; Peptide; 8 AA.  
ID AAB28957;  
AC AAB28957;  
XX 29-JAN-2001 (first entry)  
DT TRFP chain 2 peptide #1.  
DE  
XX  
XX Cat; allergy: human T cell reactive feline protein; hTRFP;  
KW immunotherapy.  
XX  
XX  
XX Felis sp.  
XX  
XX US6120769-A.  
XX 19-SEP-2000.  
XX PD 28-APR-1995; 95US-0431184.  
XX PF 02-SEP-1994; 94US-0300928.  
XX PR

Query Match 8.3%; Score 46; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 61 MKTIODCY 68  
| | | | | | | | | |  
Db 1 mktidocy 8  
RESULT 2  
AAB28957

ID AAY87698 standard; Peptide: 8 AA.

XX AAY87698;

DT 22-AUG-2000 (first entry)

DE Feline human TRFP chain 2 peptide fragment.

KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen; primer;  
KW anti-allergic; T cell stimulator; diagnostic; immunotherapy.

XX OS Felis sp.

PN US6048962-A.

PD 11-APR-2000.

PF 27-APR-1995; 95US-0430014.

PR 02-SEP-1994; 94US-0300928.

Query Match

Best Local Similarity 8.3%; Score 46; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 MKKIDCY 68  
| | | | | | | |  
Db 1 mkkidcy 8

RESULT 3

ID AAY51495 standard; Protein: 8 AA.

XX AAY51495;

DT 22-MAY-2000 (first entry)

DE Human TRFP chain 2 antigenic fragment #1.

KW T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
KW down regulation; immune response; allergen; immunoglobulin E;

KW sensitivity; cat protein allergen; human; chain 2.

XX OS Homo sapiens.

PN US6019972-A.

PD 01-FEB-2000.

PF 02-SEP-1994; 94US-0300928.

Query Match

Best Local Similarity 8.3%; Score 46; DB 21; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 61 MKKIDCY 68  
| | | | | | | |  
Db 1 mkkidcy 8

RESULT 4

ID AAB28978 standard; Peptide: 7 AA.

XX AAB28978;

DT 29-JAN-2001 (first entry)

DE Peptide #2.

KW Cat; allergy; human T cell reactive feline protein; hTRFP;

KW Immunotherapy.

XX Unidentified.

PN US6120769-A.

PD 19-SEP-2000.

PF 28-APR-1995; 95US-0431184.

PR 02-SEP-1994; 94US-0300928.

Query Match

Best Local Similarity 7.2%; Score 40; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMAETCP 27  
| | | | | | | |  
Db 1 kmaetcp 7

RESULT 5

ID AAY87723 standard; Protein: 7 AA.

XX AAY87723;

DT 22-AUG-2000 (first entry)

DE Feline human TRFP chain 2 peptide fragment.

KW T-cell reactive feline protein; TRFP; Fel d I; cat allergen;  
KW anti-allergic; T cell stimulator; diagnostic; immunotherapy.

XX OS Felis sp.

PN US6048962-A.

PD 11-APR-2000.

PF 27-APR-1995; 95US-0430014.

PR 02-SEP-1994; 94US-0300928.

Query Match

Best Local Similarity 7.2%; Score 40; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMAETCP 27  
| | | | | | | |  
Db 1 kmaetcp 7

RESULT 6

ID AAY90143 standard; Peptide: 7 AA.

XX AAY90143;

DT 13-JUL-2000 (first entry)

DE Leader sequence peptide.

KW Cat; TRFP; human T-cell reactive feline protein; cat protein allergen;  
KW house dust; Fel d I; cat allergy; Felis domesticus sensitivity; therapy;

KW diagnosis; goat; sheep; horse; rabbit; dog; leader sequence.

XX OS Unidentified.

PN US6025162-A.

PD 15-FEB-2000.

XX



DT 22-MAY-2000 (first entry)  
XX  
DE Human TRFP chain 2 antigenic fragment #3.  
XX  
XX T-cell reactive feline protein; TRFP; T cell epitope; T cell receptor;  
KW down regulation; immune response; allergen; immunoglobulin E;  
KW sensitivity; cat protein allergen; human; chain 2.  
XX  
OS Homo sapiens.  
XX  
PN US6019972-A.  
XX  
PD 01-FEB-2000.  
XX  
PF 02-SEP-1994; 94US-0300928.  
XX

Query Match 5.8%; Score 32; DB 21; Length 6;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 TEPERT 59  
111111  
Db 1 teper 6

RESULT 12  
AAG89573  
ID AAG89573 standard; Peptide; 8 AA.  
XX  
AC AAG89573;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE p53 epitope A2 supermotif crossbinding peptide #8.  
XX  
KW Human; p53; tumour suppressor; human leukocyte antigen; HLA; CTL;  
KW cytotoxic T lymphocyte; T cell; immune response; tumour; cancer;  
KW vaccine; epitope; cytostatic.  
XX  
OS Homo sapiens.  
XX  
OS Synthetic.  
XX  
PN WO200141788-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 11-DEC-2000; 2000MO-US33629;  
XX

Query Match 5.8%; Score 32; DB 22; Length 8;  
Best Local Similarity 50.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMAETCPI 28  
111111  
Db 1 qlaktcpv 8

RESULT 13  
AAR97531  
ID AAR97531 standard; peptide; 9 AA.  
XX  
AC AAR97531;  
XX  
DT 12-FEB-1997 (first entry)  
XX  
DE Antigenic peptide; corresp. to p53 aa 136-144.  
XX  
KW p53; Her-2; Neu; aa; amino acid; CTL; cytotoxic T lymphocyte; target;  
KW malignant cell; antigenic; vaccine; immunisation; activation.  
XX  
OS Homo sapiens.  
XX

PN WO9618409-A1.  
XX  
PD 20-JUN-1996.  
XX  
XX 14-DEC-1995; 95WO-US16415.  
XX  
PR 14-DEC-1994; 94US-0355558.  
XX

Query Match 5.8%; Score 32; DB 17; Length 9;  
Best Local Similarity 50.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMAETCPI 28  
111111  
Db 1 qlaktcpv 8

RESULT 14  
AAV53375  
ID AAV53375 standard; Protein; 9 AA.  
XX  
AC AAV53375;  
XX  
DT 18-JAN-2000 (first entry)  
XX  
DE p53 epitope (aa 135-143) binds HLA-B8.  
XX  
KW Lipopeptide; epitope; cytotoxic T lymphocyte; CTL; lipid; spacer; p53;  
KW electrical charge; hydrophilicity; vaccine; immune response; HIV; HBV;  
KW human immunodeficiency virus; hepatitis B virus; papilloma virus;  
KW melanoma; malaria; parasite.  
XX  
OS Synthetic.  
XX  
OS Homo sapiens.  
XX  
PN FR2776926-A1.  
XX  
PD 08-OCT-1999.  
XX

Query Match 5.8%; Score 32; DB 20; Length 9;  
Best Local Similarity 50.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 21 KMAETCPI 28  
111111  
Db 2 qlaktcpv 9

RESULT 15  
AAV40198  
ID AAV40198 standard; Peptide; 9 AA.  
XX  
AC AAV40198;  
XX  
DT 19-NOV-1999 (first entry)  
XX  
DE Amino acid sequence of a p53 protein epitope.  
XX  
KW Cytotoxic T cell; T lymphocyte; CD8+ epitope; T helper cell;  
KW CD4+ epitope; B epitope; lipopeptide; interferon gamma; adjuvant;  
KW vaccine; tumor; infection; immune response; cytokine profile;  
KW acquired immune deficiency syndrome; papilloma; cancer; hepatitis;  
KW autoimmune disease.  
XX  
OS Homo sapiens.  
XX  
PN FR2774687-A1.  
XX  
PD 13-AUG-1999.  
XX

Query Match 5.8%; Score 32; DB 20; Length 9;

Wed Jul 10 08:06:56 2002

us-09-662-784-6.rag

Page 5

Best Local Similarity 50.0%; Pred. No. 6.4e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 21 KMAETCPI 28  
: : : : : :  
Db 2 qlaktcpv 9

Search completed: July 9, 2002, 16:20:36  
Job time: 81 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:19:16 ; Search time 15.69 Seconds

(Without alignments)  
679.791 Million cell updates/sec

Title: US-09-662-784-6

Perfect score: 552  
Sequence: 1 DTMRCALVIALVLTQALGV.....CMGEAVONTVEDLKINTIGR 111

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1099

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	4.9	9	2	A91466
2	27	4.9	9	2	A92774
3	27	4.9	9	2	A93147
4	27	4.9	9	2	A93408
5	27	4.9	9	2	B90667
6	27	4.9	9	2	S06375
7	27	4.9	9	2	B61364
8	27	4.9	10	2	A61131
9	25	4.5	9	2	A61364
10	24	4.3	6	2	A19780
11	24	4.3	9	2	PS0253
12	22	4.0	6	2	I51317
13	22	4.0	9	2	PC2021
14	22	4.0	10	2	S65728
15	22	4.0	10	2	A61007
16	22	4.0	10	2	A59173
17	21	3.8	9	2	A28495
18	21	3.8	9	2	S30494
19	21	3.8	9	2	B24362
20	21	3.8	9	2	B30572
21	21	3.8	10	2	S39040
22	21	3.8	10	2	I36893
23	21	3.8	10	2	A47593
24	21	3.8	10	2	PS0220
25	21	3.8	10	2	C30572
26	20	3.6	10	2	F44644
27	20	3.6	10	2	PT0212
28	20	3.6	10	2	A55695
29	19	3.4	4	2	I40804

30	19	3.4	7	2	PT0521	T-cell receptor be
31	19	3.4	8	2	T10952	hypothetical prote
32	19	3.4	8	2	D61512	variant surface gl
33	19	3.4	8	2	A42057	fibroblast growth
34	19	3.4	8	2	PT0716	T-cell receptor be
35	19	3.4	8	2	S68325	blood cell protein
36	19	3.4	9	2	B45796	dihydroliipoamide s
37	19	3.4	9	2	S13333	alpha/beta-gliadin
38	19	3.4	10	2	G58501	48k bile/gallblad
39	19	3.4	10	2	A61622	vitellogenin, 190k
40	19	3.4	10	2	S23371	T-cell receptor al
41	18	3.3	6	2	PT0652	T-cell receptor be
42	18	3.3	7	2	S29735	polyphosphate--glu
43	18	3.3	8	2	S43971	tumor-associated a
44	18	3.3	8	2	S43972	tumor-associated a
45	18	3.3	9	2	S59902	glutathione transf

#### ALIGNMENTS

##### RESULT 1

A91466

oxytocin - hippopotamus  
N:Alternate names: neurophysin I

N:Contains: neurophysin 1; oxytocin

C:Species: Hippopotamus amphibius (hippopotamus)

C>Date: 30-Oct-1992 #sequence, revision 30-Oct-1992 #text\_change 20-Mar-1998

C:Accession: A91466; A01450; B01450

R: Ferguson, D.R.; Pickering, B.T.

Gen. Comp. Endocrinol. 13, 425-429, 1969

A:Title: Arginine and lysine vasopressins in the hippopotamus neurohypophysis.

A:Reference number: A91466; MUID:71232719

A:Accession: A91466

A:Molecule type: protein

A:Residues: 1-9 <FER>

C:Comment: Oxytocin is followed by neurophysin 1 in the precursor.

C:Keywords: amidated carboxyl end; hormone; hypothalamus

F:1-6/Dsulfide bonds: #status predicted

F:9/Modified site: amidated carboxyl end (gly) #status predicted

##### Query Match

Best Local Similarity 4.9%; Score 27; DB 2; Length 9;

Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71

Db 1 CYION 5

##### RESULT 2

A92774

oxytocin - spotted ratfish

N:Alternate names: neurophysin I

N:Contains: neurophysin 1; oxytocin

C:Species: Hydroilaqus collieri (spotted ratfish)

C>Date: 30-Oct-1992 #sequence, revision 30-Oct-1992 #text\_change 20-Mar-1998

C:Accession: A92774; A01450; B01450

R: Pickering, B.T.; Heller, H.

J. Endocrinol. 45, 597-606, 1969

A:Title: Oxytocin as a neurohypophysial hormone in the holoccephalian elasmobranch fi-

A:Reference number: A92774; MUID:70088110

A:Accession: A92774

A:Molecule type: protein

A:Residues: 1-9 <PIC>

C:Comment: Oxytocin is followed by neurophysin 1 in the precursor.

C:Superfamily: oxytocin-neurophysin

C:Keywords: amidated carboxyl end; hormone; hypothalamus

F:1-6/Dsulfide bonds: #status predicted

F:9/Modified site: amidated carboxyl end (gly) #status predicted

Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
||:|  
Db 1 CYION 5

RESULT 3  
A:Accession: A93147  
A:Title: Isolation of flinback whale oxytocin and vasopressin.  
A:Reference number: A93147  
A:Molecule type: protein  
A:Residues: 1-9 <ACH>  
C:Comment: Oxytocin is followed by neurophysin 1 in the precursor.  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: amidated carboxyl end; hormone; hypothalamus  
F:1-6/Disulfide bonds: #status experimental  
F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
||:|  
Db 1 CYION 5

RESULT 4  
A:Accession: A93408  
A:Title: Australian echidna oxytocin - flinback whale oxytocin  
N:Alternate names: neurophysin 1  
C:Species: Macroglossus aculeatus (Australian echidna)  
C:Date: 30-Oct-1992 #sequence\_revision 30-Oct-1992 #text\_change 20-Mar-1998  
C:Accession: A93408; A01450; B01450  
R:Acher, R.; Chauvet, J.; Chauvet, M.T.  
N:Gene: New Biol. 244, 124-126, 1973  
A:Title: Neurohypophysial hormones and evolution of tetrapods.  
A:Reference number: A93408; MUID:73223515  
A:Accession: A93408  
A:Molecule type: protein  
A:Residues: 1-9 <ACH>  
C:Comment: Oxytocin is followed by neurophysin 1 in the precursor.  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: amidated carboxyl end; hormone; hypothalamus  
F:1-6/Disulfide bonds: #status predicted  
F:9/Modified site: amidated carboxyl end (Gly) #status predicted

Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
||:|  
Db 1 CYION 5

RESULT 5  
1

B90667

Oxytocin - rabbit  
N:Alternate names: neurophysin 1  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 30-Oct-1992 #sequence\_revision 30-Oct-1992 #text\_change 20-Mar-1998  
C:Accession: B90667; A01450; B01450  
R:Chauvet, J.; Chauvet, M.T.; Acher, R.  
Biochimie 53, 1099-1104, 1971  
A:Title: Evolution des hormones neurohypophysiales: isolement des principes actifs du  
A:Reference number: B90667; MUID:72215060  
A:Accession: B90667  
A:Molecule type: protein  
A:Residues: 1-9 <CH>  
C:Comment: Oxytocin is followed by neurophysin 1 in the precursor.  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: amidated carboxyl end; hormone; hypothalamus  
F:1-6/Disulfide bonds: #status experimental  
F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
||:|  
Db 1 CYION 5

RESULT 6  
S06375  
arginine vasotocin - sea lamprey  
C:Species: Petromyzon marinus (sea lamprey)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 31-Mar-1997  
C:Accession: S06375  
R:Lane, T.F.; Sower, S.A.; Kawachi, H.  
Gen. Comp. Endocrinol. 70, 152-157, 1988  
A:Title: Arginine vasotocin from the pituitary gland of the lamprey (petromyzon marinus)  
A:Reference number: S06375; MUID:88225976  
A:Accession: S06375  
A:Molecule type: protein  
A:Residues: 1-9 <LAN>  
C:Keywords: amidated carboxyl end; hormone; neuropeptide; pituitary  
F:1-6/Disulfide bonds: #status experimental  
F:9/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
||:|  
Db 1 CYION 5

RESULT 7  
B61364  
vasotocin - common carp  
C:Species: Cyprinus carpio (common carp)  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Jun-2000  
C:Accession: B61364  
R:Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.  
Comp. Biochem. Physiol. A 14, 243-254, 1965  
A:Title: Caractérisation des hormones neurohypophysiales d'un poisson osseux d'eau do  
A:Reference number: B61364  
A:Accession: B61364  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-9 <ACH>  
C:Keywords: amidated carboxyl end; neuropeptide; posterior pituitary



Query Match 4.9%; Score 27; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
|||:  
Db 1 CYION 5

RESULT 8  
A61131  
hydrin 2 - bullfrog  
N:Alternate names: AVT-related peptide  
C:Species: Rana catesbeiana (bullfrog)  
C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 11-Jul-1997  
C:Accession: A61131  
R:Iwamuro, S.; Hayashi, H.; Yamashita, M.; Kikuyama, S.  
Gen. Comp. Endocrinol. 84, 412-418, 1991  
A:Title: Arginine vasotocin (AVT) and AVT-related peptide are major aldosterone-releasin  
A:Reference number: A61131; MUID:9225310  
A:Accession: A61131  
A:Molecule type: protein  
A:Residues: 1-10 <IWA>  
C:Comment: Both the decapeptide (hydrin 2) and the amidated nonapeptide (Arg-vasotocin)  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: amidated carboxyl end; neuropeptide; pituitary  
F:1-9/Product: Arg-vasotocin [status experimental <RVAS>  
F:9/Modified site: amidated carboxyl end (Gly) (amide in mature form from following gly

Query Match 4.9%; Score 27; DB 2; Length 10;  
Best Local Similarity 60.0%; Pred. No. 6.3e+03;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 67 CYVEN 71  
|||:  
Db 1 CYION 5

RESULT 9  
A61364  
isotocin - common carp  
C:Species: Cyprinus carpio (common carp)  
C:Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Jun-2000  
C:Accession: A61364  
R:Acher, R.; Chauvet, J.; Chauvet, M.T.; Crepy, D.  
Comp. Biochem. Physiol. A 14, 245-254, 1965  
A:Title: Caractérisation des hormones neurohypophysaires d'un poisson osseux d'eau douce  
A:Reference number: A61364  
A:Accession: A61364  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-9 <ACH>  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: amidated carboxyl end; neuropeptide; posterior pituitary  
F:9/Modified site: amidated carboxyl end (Gly) [status experimental

Query Match 4.5%; Score 25; DB 2; Length 9;  
Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 67 CYVEN 71  
|||:  
Db 1 CYISN 5

RESULT 10  
A19780  
transferrin - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)

C:Date: 25-Oct-1987 #sequence\_revision 25-Oct-1987 #text\_change 18-Jun-1993  
C:Accession: A19780  
R:Brock, J.H.; Esparza, I.; Oliver, R.A.; Spooner, R.L.  
Biochem. Genet. 18, 851-860, 1980  
A:Title: Electrophoretic mobility of N- and C-terminal monoferric fragments of bovin  
A:Reference number: A19780; MUID:81183891  
A:Accession: A19780  
A:Molecule type: protein  
A:Residues: 1-6 <BRO>

Query Match 4.3%; Score 24; DB 2; Length 6;  
Best Local Similarity 80.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 55 EPERT 59  
:||||  
Db 1 DPERT 5

RESULT 11  
PS0253  
glycine cleavage system protein H - rice (strain Nihonbare) (fragment)  
C:Species: Oryza sativa (rice)  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 03-Feb-1994  
C:Accession: PS0253  
R:Tsugita, A.  
submitted to JIPID, April 1993  
A:Reference number: PS0206  
A:Accession: PS0253  
A:Molecule type: protein  
A:Residues: 1-9 <TSD>

Query Match 4.3%; Score 24; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 77 VLDGL 81  
|||||  
Db 3 VLDGL 7

RESULT 12  
I51317  
bHLH transcription factor inhibitor - African clawed frog (fragment)  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: I51317  
R:Zhang, H.; Reynaud, S.; Kloc, M.; Etkin, L.D.; Spohr, G.  
Mech. Dev. 50, 119-130, 1995  
A:Title: Id gene activity during Xenopus embryogenesis.  
A:Reference number: I51316; MUID:95344988  
A:Accession: I51317  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-6 <ZHA>  
A:Cross-references: GB:S79038; NID:g1042006; PIDN:RAD14294.1; PID:g4261994  
C:Genetics:  
A:Gene: Xidib

Query Match 4.0%; Score 22; DB 2; Length 6;  
Best Local Similarity 80.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 52 NATEP 56  
:||||  
Db 1 SATEP 5

RESULT 13

PC2021  
oxytocin-related peptide, amnetocin - earthworm (Eisenia foetida)

C:Species: Eisenia foetida (common brandling worm)  
C:Date: 03-May-1994 #sequence\_revision 15-Oct-1994 #text\_change 11-Jul-1997

C:Accession: PC2021

R:Ouml, T.; Ukena, K.; Matsushita, O.; Ikeda, T.; Fujita, T.; Minakata, H.; Nomoto, K.

Biochem. Biophys. Res. Commun. 198, 393-399, 1994

A:Title: Amnetocin: an oxytocin-related peptide isolated from the earthworm, Eisenia foetida

A:Reference number: PC2021, MUID:94121660

A:Accession: PC2021

A:Molecule type: protein

A:Residues: 1-9 <OUM>

C:Comment: This protein potentiates spontaneous contractions of the gut and also pulsac

C:Keywords: amidated carboxyl end

F:1-6/Disulfide bonds: #status experimental

F:9/Modified site: amidated carboxyl end (Gly) #status experimental

#### Query Match

Best Local Similarity 4.0%; Score 22; DB 2; Length 9;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 67 CYVEN 71

DB 1 CFVRN 5

#### RESULT 14

S65728 hemoglobin, extracellular, chain d1 - earthworm (Lumbricus terrestris) (fragment)

C:Species: Lumbricus terrestris (common earthworm)

C:Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997

C:Accession: S65728

R:Fushitani, K.; Higashiyama, K.; Asao, M.; Hosokawa, K.

Biochim. Biophys. Acta 1292, 273-280, 1996

A:Title: Characterization of the constituent polypeptides of the extracellular hemoglobin

A:Reference number: S65721, MUID:96176855

A:Accession: S65728

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-10 <FUS>

#### Query Match

Best Local Similarity 4.0%; Score 22; DB 2; Length 10;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 13 LYTOALGV 20

DB 3 LVYTESLKV 10

#### RESULT 15

A61007 hementin (EC 3.4.-.-) - Amazon leech (fragment)

C:Species: Haementeria ghilianii (Amazon leech)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 07-May-1999

C:Accession: A61007

R:Swadesh, J.K.; Huang, I.Y.; Budzynski, A.Z.

J. Chromatogr. 502, 359-369, 1990

A:Title: Purification and characterization of hementin, a fibrinogenolytic protease from

A:Reference number: A61007, MUID:90256973

A:Accession: A61007

A:Molecule type: protein

A:Residues: 1-10 <SWA>

C:Keywords: anticoagulant; hydrolase; saliva

Query Match 4.0%; Score 22; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2.1e+04;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 TEPE 57

DB 4 TEPE 7

Search completed: July 9, 2002, 16:20:59  
Job time: 103 sec



---

**THIS PAGE BLANK (USPTO)**

OC Elasmobranchii; Squala; Hypnosquala; Pristiogaster; Batoidae;  
OC Rajiformes; Rajidae; Raja.  
OX NCBI\_TaxID=7781;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=66123415; PubMed=5880565;  
RA Acher R., Chauvet J., Chauvet M.-T., Crepy D.;  
RT "Phylogeny of neurohypophyseal peptides: isolation of a new hormone,  
RT glutitocin (Ser 4-gln 8-oxytocin) present in a cartilaginous fish,"  
RL the ray (Raja clavata).";  
RL Biochim. Biophys. Acta 107:393-396(1965).  
CC -!- FUNCTION: ANTIDIURETIC HORMONE.

Query Match 4.5%; Score 25; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 67 CYVEN 71  
Db 1 CYISN 5

RESULT 8  
OXYT\_BUFE STANDARD; PRT; 9 AA.  
ID OXYT\_BUFE  
AC P42995;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Seritocin.  
OS Bufo regularis (Leopard toad).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae;  
OC Bufo.  
OX NCBI\_TaxID=8390;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Pituitary neurointermediate lobe;  
RX MEDLINE=96059313; PubMed=7591488;  
RA Chauvet J., Michel G., Ouedraogo Y., Chou J., Chait B.T., Acher R.;  
RT "A new neurohypophysial peptide, seritocin (Ser5, Ile8]-oxytocin),  
RT identified in a dryness-resistant African toad, Bufo regularis.";  
RL Int. J. Pept. Protein Res. 45:482-487(1995).  
CC -!- FUNCTION: DEVOID OF OXYTOIC ACTIVITY.

Query Match 4.2%; Score 23; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 24 ETCP1 28  
Db 4 QSCPI 8

RESULT 9  
OXYT\_EISFO STANDARD; PRT; 9 AA.  
ID OXYT\_EISFO  
AC P42998;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Annelocin.  
OS Eisenia foetida (Common brandling worm) (Common dung-worm).  
OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
OC Lumbricidae; Lumbricidae; Eisenia.  
OX NCBI\_TaxID=6396;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Pituitary;  
RX MEDLINE=9412160; PubMed=8292046;  
RA Oumi T., Ukena K., Matsushima O., Ikeda T., Fujita T., Minakata H.,  
RA Nomoto K.;  
RT "Annetocin: an oxytocin-related peptide isolated from the earthworm,  
RT

RT Eisenia foetida.";  
RL Biochem. Biophys. Res. Commun. 198:393-399(1994).  
CC -!- FUNCTION: POTENTIATES SPONTANEOUS CONTRACTIONS OF THE GUT AND ALSO

Query Match 4.0%; Score 22; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 67 CYVEN 71  
Db 1 CEFVN 5

RESULT 10  
PAP1\_PARMA STANDARD; PRT; 10 AA.  
ID PAP1\_PARMA  
AC P81863;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pardaxin I (PXT) (Fragment).  
OS Pardachirus marmoratus (Red sea mores sole).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
OC Soleoidae; Soleidae; Pardachirus.  
OX NCBI\_TaxID=31087;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skin secretion;  
RX MEDLINE=87057369; PubMed=3782138;  
RA Lazarovici P., Primor N., Loew L.M.;  
RT "Purification and pore-forming activity of two hydrophobic  
RT polypeptides from the secretion of the Red sea mores sole (Pardachirus  
RT marmoratus).";

Query Match 4.0%; Score 22; DB 1; Length 10;  
Best Local Similarity 55.6%; Pred. No. 9.5e+03;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 33 FFAVANGNE 41  
Db 2 FFALIPGE 10

RESULT 11  
CONO\_CONGE STANDARD; PRT; 9 AA.  
ID CONO\_CONGE  
AC P05486;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE LYS-conopressin G.  
OS Conus geographus (Geography cone).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;  
OC Neogastropoda; Conoidea; Conidae; Conus.  
OX NCBI\_TaxID=6491;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=88058932; PubMed=3680228;  
RA Cruz L.J., de Santos V., Zafaralla G.C., Ramillo C.A., Zeikus R.D.,  
RA Gray W.R., Olivera B.M.;  
RT "Invertebrate vasopressin/oxytocin homologs. Characterization of  
RT peptides from Conus geographus and Conus stratus venoms.";  
RL J. Biol. Chem. 262:15821-15824(1987).  
RN [2]  
RP REVIEW.

Query Match 3.8%; Score 21; DB 1; Length 9;  
Best Local Similarity 40.0%; Pred. No. 1e+05; Indels 0; Gaps 0;  
Matches 2; Conservative 2; Mismatches 1;

OY 67 CYVEN 71

Db 1 CERN 5

RESULT 12  
LPCA\_STAU STANDARD; PRT: 9 AA.

AC P36884;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DE Chloramphenicol resistance leader peptide.  
OS Staphylococcus aureus, and  
OS Streptococcus agalactiae.  
OC Plasmid pSC56, plasmid pSC57, plasmid pBH12, and plasmid pIP501.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280, 1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S. aureus; STRAIN-436; PLASMID-PSC57;  
RX MEDLINE-92027653; PubMed-1929326;  
RA Schwarz S., Cardoso M.;  
RT Nucleotide sequence and phylogeny of a chloramphenicol  
RT acetyltransferase encoded by the plasmid pSC57 from Staphylococcus  
RT aureus.";

Query Match  
Best Local Similarity 3.8%; Score 21; DB 1; Length 9;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 61 MKKIOD 66  
Db 1 MKKSD 6

RESULT 13

OXYF\_SCYCA

STANDARD; PRT: 9 AA.

AC P42997;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Phasvatocin.  
OS Scyllorhinus canicula (Spotted dogfish) (Spotted catshark).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Galeomorphi; Galeidae; Carcharhiniformes;  
OC Scyllorhinidae; Scyllorhinus.  
OX NCBI\_TaxID=7830;  
RN [1]  
RP SEQUENCE.  
RC TISSUE-Pituitary;  
RX MEDLINE-95062247; PubMed-7972045;  
RA Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;  
RT \*Special evolution of neurohypophyseal hormones in cartilaginous  
RT fishes: asvatocin and phasvatocin, two oxytocin-like peptides  
RT isolated from the spotted dogfish (Scyllorhinus canicula).";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).

Query Match  
Best Local Similarity 3.8%; Score 21; DB 1; Length 9;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 67 CYVEN 71  
Db 1 CYENN 5

RESULT 14

OXYT\_OCTVU

ID OXYT\_OCTVU

STANDARD;

PRT:

9 AA.

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Cephalotocin.  
OS Octopus vulgaris (Octopus).  
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Octopoda;  
OC Incirrata; Octopodidae; Octopus.  
OX NCBI\_TaxID=6645;  
RN [1]  
RP SEQUENCE.  
RC TISSUE-Nerve endings;  
RX MEDLINE-92270139; PubMed-1589145;  
RA Reich G.;  
RT "A new peptide of the oxytocin/vasopressin family isolated from  
RT nerves of the cephalopod Octopus vulgaris.";  
RL Neurosci. Lett. 134:191-194(1992).  
CC -1- FUNCTION: HAS A ROLE IN THE NEUROSECRETORY SYSTEM OF THE VENA  
CC CAVA.

Query Match  
Best Local Similarity 3.8%; Score 21; DB 1; Length 9;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 67 CYVEN 71  
Db 1 CYERN 5

RESULT 15

XYNB\_DICB4

STANDARD; PRT: 10 AA.

AC P80717;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Endo-1,4-beta-xylanase B (EC 3.2.1.8) (Xylanase B)  
DE (1,4-beta-D-xylan xylanohydrolase B) (fragment).  
OS Dictyoglomus sp. (Strain B4A).  
OC Bacteria; Dictyoglomus group; Dictyoglomus.  
OX NCBI\_TaxID=69007;  
RN [1]  
RP SEQUENCE.  
RA Adamsen A.K., Jacobsen S., Ahning B.K.;  
RL Submitted (Oct-1996) to the SWISS-PROT data bank.  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic  
CC linkages in xyans.  
CC -1- PATHWAY: XYLAN DEGRADATION.  
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL  
CC HYDROLASES).  
DR InterPro; IPR001000; Glyco\_hydro.10.

Query Match  
Best Local Similarity 3.8%; Score 21; DB 1; Length 10;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 98 ONTVEDLK 105  
Db 2 KTIIDLK 9

Search completed: July 9, 2002, 16:21.48  
Job time: 117 sec



**THIS PAGE BLANK (USPTO)**

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:19:51 ; Search time 11.72 Seconds

(without alignments)  
366.712 Million cell updates/sec

Title: US-09-662-784-6  
Perfect score: 552  
Sequence: 1 DFWRGALVLLVLTQALGV.....CMGEAVONTVEKLNTLGR 111

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 349

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	27	4.9	9	1 OXYT_CYPCA	P23879 cyprinus ca
2	27	4.9	9	1 OXYT_RABIT	P32878 oryctolagus
3	27	4.9	9	1 OXYV_SQUAC	P43000 squallus aca
4	25	4.5	9	1 ISOT_CYPCA	P42993 cyprinus ca
5	25	4.5	9	1 OXYA_SCYCA	P42996 scyllorhinu
6	25	4.5	9	1 OXYA_SQUAC	P42999 squallus aca
7	25	4.5	9	1 OXYT_RAUCI	P42994 raja clavata
8	23	4.2	9	1 OXYT_EISFO	P42995 buto regula
9	22	4.0	9	1 OXYT_EISFO	P42998 eisenia foe
10	22	4.0	10	1 CONO_CONGE	P05486 conus geogr
11	21	3.8	9	1 LPCA_STANU	P32884 staphylococ
12	21	3.8	9	1 OXYF_SCYCA	P42997 scyllorhinu
13	21	3.8	9	1 OXYT_OCTUV	P80027 octopus vul
14	21	3.8	10	1 XYNB_DICB4	P80717 dictyoglomu
15	21	3.6	9	1 FLA2_TREHY	P80159 treponema h
16	20	3.6	5	1 UC22_MAITE	P80628 zea mays (m
17	19	3.4	9	1 FAR5_PANRE	P82661 panagrellus
18	19	3.4	10	1 PCK_FASHE	P80525 fasciola he
19	18	3.3	8	1 CAD1_ENTFA	P13268 enterococu
20	18	3.3	10	1 RL16_ACHLA	P23221 acholeplasm
21	18	3.3	10	1 UP11_CAEEL	P55956 caenorhabdi
22	18	3.3	7	1 CTA_ENTFA	P11932 enterococu
23	17	3.1	7	1 E105_LITRU	P81201 littoria rub
24	17	3.1	7	1 IGDO_DACDE	P05487 dactylium d
25	17	3.1	9	1 CONO_CONST	P05487 conus stria
26	17	3.1	9	1 YBR3_CAIYO	P24858 calliphora
27	17	3.1	9	1 YBR3_AZOVI	P24858 azotobacter
28	17	3.1	10	1 COXQ_RABIT	P80336 oryctolagus
29	17	3.1	10	1 GAUV_HUMAN	P01358 homo sapien
30	17	3.1	10	1 PSBF_CAPAN	P003367 capisicum an
31	17	3.1	10	1 TKN1_SCYCA	P08608 scyllorhinu
32	17	3.1	10	1 UHA3_HUMAN	P40930 homo sapien
33	17	3.1	10	1	

34	17	3.1	10	1	UPA4_HUMAN	P30090 homo sapien
35	16	2.9	5	1	RE11_LITRU	P82070 littoria rub
36	16	2.9	5	1	RE21_LITRU	P82071 littoria rub
37	16	2.9	6	1	UN06_CLOPA	P81351 clostridium
38	16	2.9	8	1	NPB_BOVIN	P15507 bos taurus
39	16	2.9	9	1	COVN_CONVE	P83047 conus ventr
40	16	2.9	9	1	DNF1_LOCOM	P16339 locusta mig
41	16	2.9	9	1	DSIP_RABIT	P01158 oryctolagus
42	16	2.9	9	1	FARA_CAIYO	P41865 calliphora
43	16	2.9	9	1	RE42_LITRU	P82075 littoria rub
44	16	2.9	9	1	RE43_LITRU	P82093 littoria rub
45	16	2.9	9	1	SAP_STOVA	P24047 stomopneute

## ALIGNMENTS

RESULT 1  
ID OXYT\_CYPCA STANDARD; PRT; 9 AA.  
AC P23879:  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Vasotocin.  
OS Cyprinus carpio (Common carp), and  
OS Petromyzon marinus (Sea lamprey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Cyprinus.  
OX NCBI\_TaxID=7962, 7757;  
RN [1]  
RP SEQUENCE.  
RC SPECIES=C. carpio; TISSUE=Pituitary;  
RA Acher R., Chauvet J., Chauvet M.-T., Crepy D.;  
RT "Characterization of neurohypophyseal hormones from a fresh water bony  
RT fish, the carp (Cyprinus carpio). Comparison with hormones from sea  
RT water bony fishes."  
RL Comp. Biochem. Physiol. 14:245-254(1965).

Query Match 4.9%; Score 27; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 67 CYVEN 71  
DB 1 CYION 5

RESULT 2  
ID OXYT\_RABIT STANDARD; PRT; 9 AA.  
AC P32878; P01188;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Oxytocin (Oxytocin).  
OS Oryctolagus cuniculus (Rabbit);  
OS Hippopotamus amphibius (Hippopotamus);  
OS Balanoptera physalus (Finback whale) (Common rorqual);  
OS Tachygylossus aculeatus aculeatus (Austrian eel);  
OS Hydrolyagus colliei (Spotted ratfish) (Pacific ratfish);  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986, 9833, 9770, 49271, 7873;  
RN [1]  
RP SEQUENCE.  
RC SPECIES=Rabbit;  
RA MEDLINE=72215060; Pubmed=5150741;  
RA Chauvet J., Chauvet M.-T., Acher R.;  
RT "Evolution of neurohypophyseal hormones: Isolation of active

Query Match 4.9%; Score 27; DB 1; Length 9;

Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 67 CYEN 71  
DB 1 CYON 5

RESULT 3  
ID OXYV\_SQVAC STANDARD; PRT; 9 AA.  
AC P43000;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Valtocin.  
OS Squalus acanthias (Spiny dogfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squalae; Squaloidei; Squalidae; Squalus.  
OX NCBI\_TaxID=7797;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=73031727; PubMed=5083097;  
RA Acher R., Chauvet J., Chauvet M.-T.;  
RT "Phylogeny of the neurohypophysial hormones. Two new active peptides isolated from a cartilaginous fish, Squalus acanthias.";  
RL Eur. J. Biochem. 29:12-19(1972).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=72128038; PubMed=4622083;

Query Match 4.9%; Score 27; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 67 CYEN 71  
DB 1 CYON 5

RESULT 4  
ID ISOT\_CYPCA STANDARD; PRT; 9 AA.  
AC P42993;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Isotocin.  
OS Cyprinus carpio (Common carp).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Cyprinus.  
OX NCBI\_TaxID=7962;  
RN [1]  
RP SEQUENCE.  
RX TISSUE=pituitary;  
RA Acher R., Chauvet J., Chauvet M.-T., Crey D.;  
RT "Characterization of neurohypophysial hormones from a fresh water bony fish, the carp (Cyprinus carpio). Comparison with hormones from sea water bony fishes."  
RL Comp. Biochem. Physiol. 14:245-254(1965).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=72128038; PubMed=4622083;

Query Match 4.5%; Score 25; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 67 CYEN 71  
DB 1 CYISN 5

RESULT 5

OXYA\_SCYCA STANDARD; PRT; 9 AA.  
ID OXYA\_SCYCA  
AC P42996;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Asvatocin.  
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;  
OC Scyliorhinidae; Scyliorhinus.  
OX NCBI\_TaxID=7830;  
RN [1]  
RP SEQUENCE.  
RX TISSUE=pituitary;  
RX MEDLINE=95062247; PubMed=7972045;  
RA Chauvet J., Rouille F., Chauveau C., Chauvet M.-T., Acher R.;  
RT "Special evolution of neurohypophysial hormones in cartilaginous fishes: asvatocin and phasvatocin, two oxytocin-like peptides isolated from the spotted dogfish (Scyliorhinus caniculus).";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).

Query Match 4.5%; Score 25; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 67 CYEN 71  
DB 1 CYINN 5

RESULT 6  
ID OXYA\_SQVAC STANDARD; PRT; 9 AA.  
AC P42993;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Aspartocin (Aspartocin).  
OS Squalus acanthias (Spiny dogfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squalae; Squaloidei; Squalidae; Squalus.  
OX NCBI\_TaxID=7797;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=73031727; PubMed=5083097;  
RA Acher R., Chauvet J., Chauvet M.-T.;  
RT "Phylogeny of the neurohypophysial hormones. Two new active peptides isolated from a cartilaginous fish, Squalus acanthias.";  
RL Eur. J. Biochem. 29:12-19(1972).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=72128038; PubMed=4622083;

Query Match 4.5%; Score 25; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1e+05;  
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 67 CYEN 71  
DB 1 CYINN 5

RESULT 7  
ID OXYT\_RAJCL STANDARD; PRT; 9 AA.  
AC P42994;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Glumitocin.  
OS Raja clavata (Thornback ray).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:19:16 ; Search time 24.33 seconds

(without alignments)  
789.249 Million cell updates/sec

Title: US-09-662-784-6

Perfect score: 552

Sequence: 1 DTMGALVLLVLTQALGV.....CMGEAVQNTVEDLKNTIGR 111

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 1088

Minimum DB seq length: 0

Maximum DB seq length: 10

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	4.5	10	10	P82937
2	23	4.2	8	2	Q9ZIE9
3	23	4.2	10	4	Q9UMK9
4	23	4.2	10	11	O9QVF7
5	22	4.0	8	2	O93SP2
6	22	4.0	10	2	O9R5M4
7	22	4.0	10	12	O39949
8	22	4.0	10	12	O9W910
9	21	3.8	9	2	O9R7E8
10	21	3.8	10	5	O27396
11	21	3.8	10	5	O95NM8
12	21	3.8	10	12	O39948
13	21	3.8	10	12	O9W8B6
14	20	3.6	7	2	P70804
15	20	3.6	7	11	O55184
16	20	3.6	9	2	O9R370

17	20	3.6	9	2	Q9R792	Q9R792 borrelia bu
18	20	3.6	9	2	O31363	O31363 borrelia ga
19	20	3.6	9	4	P78484	P78484 homo sapien
20	20	3.6	9	10	P82440	P82440 nicotiana t
21	20	3.6	10	2	Q9R5N6	Q9R5N6 clostridium
22	20	3.6	10	2	O9S3J6	O9S3J6 escherichia
23	20	3.6	10	2	O9R791	O9R791 borrelia af
24	20	3.6	10	4	O9H3R9	O9H3R9 homo sapien
25	20	3.6	10	5	O25355	O25355 locusta mig
26	20	3.6	10	6	O9R47	O9R47 bos taurus
27	20	3.6	10	6	O9SM70	O9SM70 trichosurus
28	20	3.6	10	8	O96041	O96041 oenothera b
29	19	3.4	8	2	O9RO57	O9RO57 buchnera ap
30	19	3.4	8	2	O9R049	O9R049 buchnera ap
31	19	3.4	8	2	P72221	P72221 pseudomonas
32	19	3.4	8	4	O9Y4U3	O9Y4U3 homo sapien
33	19	3.4	8	5	O15899	O15899 babesia ovi
34	19	3.4	8	5	O19956	O19956 gossypium a
35	19	3.4	8	8	O19958	O19958 gossypium b
36	19	3.4	8	8	O19959	O19959 gossypium t
37	19	3.4	8	8	O19960	O19960 gossypium m
38	19	3.4	8	8	O19961	O19961 gossypium d
39	19	3.4	8	13	P82083	P82083 limodryast
40	19	3.4	9	2	O51765	O51765 pseudomonas
41	19	3.4	9	6	O9XSL0	O9XSL0 capra hircu
42	19	3.4	9	12	O88953	O88953 vaccinia vi
43	19	3.4	9	15	O85599	O85599 motoney mur
44	19	3.4	10	4	O60912	O60912 homo sapien
45	19	3.4	10	5	P82222	P82222 bombyx mori

#### ALIGNMENTS

RESULT 1  
P82937 PRELIMINARY; PRT; 10 AA.  
AC P82937;  
DT 01-MAR-2001 (TREMUREL. 16, Created)  
DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)  
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)  
DE UNKNOWN ENDOSPERM PROTEIN B (FRAGMENT).  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;  
OC Trilicaceae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=CV. BOWI. TISSUE=SPARCHY ENDOSPERM;  
RX MEDLINE=21088911; PubMed=11271488;  
RA Kristoffersen H.E., Flengsrud R.;  
RT "Separation and characterization of basic barley seed proteins.";  
RL Electrophoresis 21:3693-3700(2000).  
CC -I- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
CC PROTEIN IS: 8.5-9.0, ITS MW IS: 11.9 KDA.

Query Match 4.5%; Score 25; DB 10; Length 10;  
Best Local Similarity 66.7%; Pred. No. 2.2e+04;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 65 ODCYVE 70  
DB 2 KDCYHE 7  
RESULT 2  
ID Q9ZIE9 PRELIMINARY; PRT; 8 AA.  
AC Q9ZIE9;  
DT 01-MAY-1999 (TREMUREL. 10, Created)  
DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)  
DT 01-MAY-1999 (TREMUREL. 10, Last annotation update)

DE CARBAMOYL-PHOSPHATE SYNTHASE SUBUNIT B (FRAGMENT).  
 GN CARB.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1527;  
 RX MEDLINE=95291461; PubMed=7773412;  
 RA Lawson F.S., Billows F.M., Dillon J.A.;  
 RT "Organization of carbamoyl-phosphate synthase genes in Neisseria gonorrhoeae includes a large, variable intergenic sequence which is also present in other Neisseria species.";  
 RL Microbiology 141:0-0(0).  
 RN [2]

Query Match  
 Best Local Similarity 57.18; Score 23; DB 2; Length 8;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 56 PERTAMK 62  
 Db 2 PKRTDLK 8

RESULT 3  
 ID O9UMK9 PRELIMINARY; PRT; 10 AA.  
 AC O9UMK9;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE DYSTROPHIN (FRAGMENT).  
 GN DAG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OX Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RP [1]  
 RA Saad F.A., Moastacciolo M.L., Merini L., Trevisan C., Tomelleri G.,  
 RA Angelini C., Danieli G.A.;  
 RT "Novel point mutations in the human dystrophin gene detected by double strand conformation analysis";  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L41643; AAB59464.1;  
 FT NON\_TER 1

Query Match  
 Best Local Similarity 80.08; Score 23; DB 4; Length 10;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 62 KRIOD 66  
 Db 1 KRIOD 5

RESULT 4  
 ID O9OV7 PRELIMINARY; PRT; 10 AA.  
 AC O9OV7;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE BETA 2-GLYCOPROTEIN 1, BETA 2-GPI.  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN NCBI\_TaxID=10118;  
 RP SEQUENCE.  
 RX MEDLINE=92135065; PubMed=1777418;  
 RA Matsura E., Igarashi M., Igarashi Y., Nagae H., Ichikawa K.,

RA Yasuda T., Koike T.;  
 RT "Molecular definition of human beta 2-glycoprotein 1 (beta 2-GPI) by cDNA cloning and inter-species differences of beta 2-GPI in alternation of anticardiolipin binding.";  
 RL Int. Immunol. 3:1217-1221(1991).  
 SQ SEQUENCE 10 AA; 1100 MW; 94E681B767376EAL CRC64;

Query Match  
 Best Local Similarity 75.08; Score 23; DB 11; Length 10;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 25 TCPL 28  
 Db 3 TCPL 6

RESULT 5  
 ID O93SP2 PRELIMINARY; PRT; 8 AA.  
 AC O93SP2;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE QUATERNARY AMMONIUM COMPOUND-RESISTANCE PROTEIN QACDELTA1 (FRAGMENT).  
 GN QACDELTA1.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OX Pseudomonas.  
 RN NCBI\_TaxID=287;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN-FMC704; TRANSPOSON-CLASS I, INTEGRON.  
 RA Lee K., Chong Y., Yum J.H., Yong D., Livermore D.M.;  
 RT "YIM-2 metallo-beta-lactamase gene-containing integron in a Pseudomonas aeruginosa clinical isolate.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY029772; AAK50441.1;

Query Match  
 Best Local Similarity 50.08; Score 22; DB 2; Length 8;  
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 3 MRCALVLT 10  
 Db 1 MKGVLEFL 8

RESULT 6  
 ID O9RSN4 PRELIMINARY; PRT; 10 AA.  
 AC O9RSN4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE NEUROTOXIN TYPE A HN+ 35 KDA SUBUNIT.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OX Clostridium.  
 RN NCBI\_TaxID=1491;  
 RP [1]  
 RA SEQUENCE.  
 RX MEDLINE=92143938; PubMed=1781887;  
 RA Somers E., Das Gupta B.R.;  
 RT "Clostridium botulinum types A, B, C1, and E produce proteins with or without hemagglutinating activity: do they share common amino acid sequences and genes?";  
 RL J. Protein Chem. 10:415-425(1991).  
 SQ SEQUENCE 10 AA; 1143 MW; CFC68A44735B456 CRC64;

Query Match  
 Best Local Similarity 42.98; Score 22; DB 2; Length 10;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 97 VONTVED 103  
:|:|:|  
Db 2 IONSLND 8

RESULT 7  
039949  
ID 039949 PRELIMINARY; PRT; 10 AA.  
AC 039949; 039950; 039953; 039954; 039955;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE E1 PROTEIN (FRAGMENT).  
OS Hepatitis GB virus C.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC GBV-C/HGV group.  
OX NCBI\_TaxID=39839;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ED\_INBURNCH HAEMOPHILIC;  
RX MEDLINE=97368412; PubMed=9225026;  
RA Smith D.B., Cuccaneu N., Davidson F., Jarvis L.M., Mokili J., Hamid S.,  
RA Ludlam C.A., Simmonds P.;  
RT "Discrimination of hepatitis G virus/GBV-C geographical variants by  
RT analysis of the 5' non-coding region."  
RL J. Gen. Virol. 78:1533-1542(1997).  
DR EMBL: AF003167; AAC57978.1; -.

Query Match  
Best Local Similarity 4.0%; Score 22; DB 12; Length 10;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 ALLVALLV 14  
|:|:|:|  
Db 2 AVLLLLLV 10

RESULT 8  
09W910  
ID 09W910 PRELIMINARY; PRT; 10 AA.  
AC 09W910;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCB-2001 (TREMBLrel. 18, Last annotation update)  
DE POLYPROTEIN (FRAGMENT).  
OS Hepatitis G virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC GBV-C/HGV group.  
OX NCBI\_TaxID=45255;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SG3316, SG57, AND SG72;  
RA Wong J., Chan S.H., Ren E.C.;  
RT "HGV isolates from Singapore: Evidence for novel Asian variants."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF078054; AAC32361.1; -  
DR EMBL: AF078048; AAC32358.1; -  
DR EMBL: AF078049; AAC32359.1; -  
KW Polyprotein.

Query Match  
Best Local Similarity 4.0%; Score 22; DB 12; Length 10;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 ALLVALLV 14  
|:|:|:|  
Db 2 AVLLLLLV 10

RESULT 9  
Q9R7E8

ID Q9R7E8 PRELIMINARY; PRT; 9 AA.  
AC Q9R7E8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE KPSD PROTEIN (FRAGMENT).  
GN KPSD.  
OS Escherichia coli.  
OC Plasmid PCR3.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95180691; PubMed=7875563;  
RA Rosenow C., Roberts I.S., Jann K.;  
RT "Isolation from recombinant Escherichia coli and characterization of  
RT CMP-Kdo synthetase, involved in the expression of the capsular K5  
RT polysaccharide (K-CKS)."  
RL FEMS Microbiol. Lett. 125:159-164(1995).

Query Match  
Best Local Similarity 3.8%; Score 21; DB 2; Length 9;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 GALLVLL 12  
|:|:|:|  
Db 2 GAKVLLSL 9

RESULT 10  
027396  
ID 027396 PRELIMINARY; PRT; 9 AA.  
AC 027396;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RHOPTRY ASSOCIATED PROTEIN 1.  
GN RAP-1.  
OS Babesia bovis.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5865;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-M07.  
RA Stares C.E., Palmer G.H., Hines S.A., McElwain T.F.;  
RT "The Babesia bovis rhoptry associated protein-1 intergenic region  
RT encodes a functional eukaryotic promoter."  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U77326; AAA96415.1; -  
SQ SEQUENCE 9 AA; 931 MW; 6D5BD2C6E5B05044 CRC64;

Query Match  
Best Local Similarity 3.8%; Score 21; DB 5; Length 9;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 76 RVLDGLV 82  
|:|:|:|  
Db 2 RIISGVV 8

RESULT 11  
095NT8  
ID 095NT8 PRELIMINARY; PRT; 10 AA.  
AC 095NT8;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CYTOCHROME P450 (FRAGMENT).  
GN CYP6D3.  
OS Musca domestica (House fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Muscoidea; Muscidae; Musca  
 OK NCBI\_TaxID=7370;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LPR  
 RA Kralj S., Scott J.G.;  
 RT #57 flanking sequence of CYP6D3.  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

Query Match  
 Best Local Similarity 3.8%; Score 21; DB 5; Length 10;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 8 ALVALLIV 15  
 Db 1 MULTIFLIT 8

RESULT 12  
 039948  
 ID 039948 PRELIMINARY; PRT; 10 AA.  
 AC 039948: 039956: 039959: Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE EL PROTEIN (FRAGMENT).  
 OS Hepatitis GB virus C.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC GBV-C/HGV group 1;  
 OK NCBI\_TaxID=39839;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ED INBURCH HAWMOHILLAC;  
 RX MEDLINE=97368412; PubMed=9225026;  
 RA Smith D.B., Cuccione N., Davidson F., Jarvis L.M., Mokili J., Hamid S.,  
 RA Ludlam C.A., Simmons P.,  
 RT Discrimination of hepatitis G virus/GBV-C geographical variants by  
 RT analysis of the 5' non-coding region."  
 RL J. Gen. Virol. 78:1533-1542(1997).  
 DR EMBL; AF003166; AAC5797.1; -.

Query Match  
 Best Local Similarity 3.8%; Score 21; DB 12; Length 10;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 ALVALLIV 14  
 Db 2 AVLLFLV 10

RESULT 13  
 09W8B6  
 ID 09W8B6 PRELIMINARY; PRT; 10 AA.  
 AC 09W8B6:  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE POLYPROTEIN (FRAGMENT).  
 OS Hepatitis G virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC GBV-C/HGV group 1;  
 OK NCBI\_TaxID=45255;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SG3410 AND SG3401;  
 RA Wong J., Chan S.H., Ren E.C.;  
 RT "HCV isolates from Singapore: Evidence for novel Asian variants."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF078064; AAC32370.1;  
 DR EMBL; AF078062; AAC32368.1; -.

KW Polyprotein. 10 10  
 FT NON\_TER  
 Query Match  
 Best Local Similarity 3.8%; Score 21; DB 12; Length 10;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 ALVALLIV 14  
 Db 2 AVLLFLV 10

RESULT 14  
 P70804  
 ID P70804 PRELIMINARY; PRT; 7 AA.  
 AC P70804:  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALGT PROTEIN (FRAGMENT).  
 GN ALGT.  
 OS Azotobacter vinelandii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Azotobacter.  
 OK NCBI\_TaxID=354;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-E;  
 RX MEDLINE=96427318; PubMed=8830682;  
 RA Rehm B.H.A., Ertsevag H., Valla S.;  
 RT "A new Azotobacter vinelandii mannuronan C-5-epimerase gene (algC) is  
 RT part of an alg gene cluster physically organized in a manner similar  
 RT to that in Pseudomonas aeruginosa."  
 RL J. Bacteriol. 178:5884-5889(1996).

Query Match  
 Best Local Similarity 3.6%; Score 20; DB 2; Length 7;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 85 TISSS 89  
 Db 2 TVSSS 6

RESULT 15  
 055184  
 ID 055184 PRELIMINARY; PRT; 7 AA.  
 AC 055184:  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ORPHAN RECEPTOR TR4-NS (FRAGMENT).  
 GN TR4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OK NCBI\_TaxID=10116;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RX MEDLINE=96198747; PubMed=8612486;  
 RA Yoshikawa T., Makino S., Gao X.M., Xing G.O., Chuang D.M.,  
 RA Defera-Wadleigh S.D.;  
 RT "Splice variants of rat TR4 orphan receptor: differential expression  
 RT of novel sequences in the 5'-untranslated region and C-terminal  
 RT domain."

Query Match  
 Best Local Similarity 3.6%; Score 20; DB 11; Length 7;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 90 KDCWGEA 96  
 I I I I I

Wed Jul 10 08:06:58 2002

us-09-662-784-6.rpt

Page 5

Db 1 KICGDA 7

Search completed: July 9, 2002, 16:21:31  
Job time: 135 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:18:10 ; Search time 11.85 seconds

(without alignments)  
362.689 million cell updates/sec

Title: US-09-662-784-6

Perfect score: 111

Sequence: 1 DTMGALIVLALVTQALGV.....CMGEAVONTVEDLKLTIGR 111

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	98.2	109	1 FEL2_FELCA	P30440 felis silve
2	8	7.2	139	1 LAMP_PETMA	P33575 petromyzon
3	8	7.2	299	1 YCXC_CYAPA	P31606 cyanophora
4	8	7.2	367	1 Y4HA_RHISN	P55471 rhizobium s
5	8	7.2	378	1 NAGA_VIBCH	O32445 vibrio chol
6	8	7.2	1106	1 ITA7_RAT	O63258 rattus norv
7	8	7.2	1179	1 ITA7_MOUSE	O61738 mus musculu
8	8	7.2	1181	1 ITA7_HUMAN	O13683 homo sapien
9	8	7.2	141	1 YEDD_SALTY	O06399 salmonella
10	7	6.3	146	1 LYC2_PIG	P12068 sus scrofa
11	7	6.3	148	1 LYC3_PIG	P12069 sus scrofa
12	7	6.3	250	1 BACR_HAISA	O93740 halobacteri
13	7	6.3	322	1 COX2_BACSU	P34957 bacillus su
14	7	6.3	332	1 C1B2_CAVPO	O94221 cavia porce
15	7	6.3	337	1 YE71_HAEIN	O57130 haemophilus
16	7	6.3	397	1 UL50_HCMVA	P16791 human cytom
17	7	6.3	400	1 ARGD_MYCTU	P94990 mycobacteri
18	7	6.3	420	1 HIS2_SYMP7	O55257 synechococc
19	7	6.3	424	1 ODB2_BACSU	P37942 bacillus su
20	7	6.3	434	1 GSA_CERSY	O74038 cenarchaeum
21	7	6.3	531	1 R060_DEIRA	O91973 delinococcus
22	7	6.3	574	1 ILVI_ECOLI	P00893 escherichia
23	7	6.3	574	1 ILVI_SALTY	P40811 salmonella
24	7	6.3	707	1 MM09_RABIT	P41246 cryptolagus
25	7	6.3	712	1 MM09_BOVIN	P52126 bos taurus
26	7	6.3	729	1 CUL6_CAEEL	O21346 caenorhabdi
27	7	6.3	899	1 PR06_YEAST	P19735 saccharomyc
28	7	6.3	1189	1 ITAH_HUMAN	O94455 homo sapien
29	7	6.3	1345	1 VCAP_HSV6U	P17887 human herpe
30	7	6.3	1345	1 POLG_TBREV	P07720 t genome po
31	7	6.3	3414	1 POLG_TBREV	O01299 t genome po
32	7	6.3	3414	1 POLG_TBREV	P14336 t genome po
33	6	5.4	80	1 ITF_HUMAN	O07654 homo sapien

34	6	5.4	100	1	SECG_AQUAE	O66505 aquifex ae
35	6	5.4	101	1	VNST_BUNYM	P16494 bunyamwera
36	6	5.4	101	1	VNST_MAGV	P16605 maguari vir
37	6	5.4	106	1	KACA_RAT	P01836 rattus norv
38	6	5.4	108	1	INS_BRARE	O73727 brachydanio
39	6	5.4	114	1	DCHS_LACBU	P04193 lactobacill
40	6	5.4	115	1	MERT_SHEPU	O54462 shewanella
41	6	5.4	120	1	CHH4_PENMO	O97386 penaeus mon
42	6	5.4	123	1	YBAV_ECOLI	P77415 escherichia
43	6	5.4	125	1	RNP_SPAEH	P16414 spaltax leuc
44	6	5.4	127	1	CHMO_BACSU	P19080 bacillus su
45	6	5.4	128	1	Y070_TREPA	O83109 treponema p

## ALIGNMENTS

RESULT 1  
ID FEL2\_FELCA STANDARD: PRT: 109 AA.  
AC P30440:  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Major allergen I polypeptide chain 2 precursor (Allergen Fel d 1-B)  
DE (Fel d 1-B) (Allergen Cat-1) (Ag4) (Fdl).  
CN CH2:  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
OX NCBI\_TaxID=9685;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-100.  
RX MEDLINE=92052157; PubMed=1946388;  
RA Morgenstern J.P., Griffith I.J., Brauer A.W., Rogers B.L.,  
RA Bond J.F., Chapman M.D., Kuo M.-C.;  
RT "Amino acid sequence of Fel d1, the major allergen of the domestic  
RT cat: protein sequence analysis and cDNA cloning.";  
RT Proc. Natl. Acad. Sci. U.S.A. 88:9690-9694(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Liver;  
RA MEDLINE=92241678; PubMed=1572548;  
RA Griffith I.J., Craig S., Pollock J., Yu X.-B., Morgenstern J.P.,  
RA Rogers B.L.;  
RT "Expression and genomic structure of the genes encoding Fdl, the  
RT major allergen from the domestic cat.";  
RT Gene 113:263-268(1992).  
RN [3]  
RP SEQUENCE OF 18-37, AND CHARACTERIZATION.  
RX MEDLINE=91287714; PubMed=1712068;  
RA Diffort O.A., Carreira J., Nitti G., Polo F., Lombardero M.;  
RT "Studies on the biochemical structure of the major cat allergen Felis  
RT domesticus I.";  
RT Mol. Immunol. 28:301-309(1991).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=84255679; PubMed=6747135;  
RA Lettermann K., Ohman J.L. Jr.;  
RT "Cat allergen I: Biochemical, antigenic, and allergenic properties.";  
RT J. Allergy Clin. Immunol. 74:147-153(1984).  
CC -1- SUBUNIT: HETEROETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED  
CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.  
CC -1- ALTERNATIVE PRODUCTS: THE LONG (CH2L) AND THE SHORT (CH2S) FORMS  
CC MAY ARISE BY ALTERNATIVE SPLICING OR MAY REPRESENT DIFFERENT  
CC ALLELES OF THE CH2 GENE. THE SEQUENCE SHOWN HERE IS THAT OF THE  
CC LONG FORM (CH2L).  
CC -1- TISSUE SPECIFICITY: THE LONG FORM IS PREFERENTIALLY EXPRESSED IN  
CC THE SALIVARY GLAND, WHILE THE SHORT FORM IS PREFERENTIALLY  
CC EXPRESSED IN THE SKIN.  
CC -1- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77341: AAC41616.1;  
DR EMBL: X62478: CAA4345.1;  
DR PIR: J01127: J01127;  
DR PIR: J01145: J01145;  
KW Allergen; Glycoprotein; Signal; Polymorphism; Alternative splicing.  
FT SIGNAL 1 17  
FT CHAIN 1 109  
FT CARBOHYD 50 50  
FT VARSPPLIC 82 82  
FT VARSPPLIC 83 109  
FT VARIANT 72 72  
FT VARIANT 72 72  
FT VARIANT 74 75  
FT VARIANT 82 83  
FT VARIANT 85 85  
FT VARIANT 86 86  
FT VARIANT 87 88  
FT VARIANT 89 89  
FT VARIANT 96 96  
FT VARIANT 105 105  
FT CONFLICT 24 24  
FT CONFLICT 32 32  
FT SEQUENCE 109 AA: 11854 MW: 857FB9CD/6036CB9 CRC64;  
  
Query Match 98.2%; Score 109; DB 1; Length 109;  
Best Local Similarity 100.0%; Pred. No. 3.8e-100;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 MRGALLVLTALVTQALGVKMAETCPFFDYVFAVANENELLDLSITVKNTEPERPAMK 62  
DB 1 MRGALLVLTALVTQALGVKMAETCPFFDYVFAVANENELLDLSITVKNTEPERPAMK 60  
QY 63 KIQDCYVNGLSIVLDGLVMTTSSSKDCKGAEVONTVEDLKNLTGR 111  
DB 61 KIQDCYVNGLSIVLDGLVMTTSSSKDCKGAEVONTVEDLKNLTGR 109  
  
RESULT 2  
LAMP\_PETMA STANDARD; PRT; 139 AA.  
ID LAMP\_PETMA  
AC P33575: P33576;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Lampyrin 0.9 precursor (Cartilage matrix protein).  
OS Petromyzon marinus (Sea lamprey).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Hyperoartia;  
OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
OX NCBI\_TaxID=7757;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 20-44.  
RC TISSUE=Cartilage; PubMed=7678258;  
RA Robson P., Wright G.M., Stelzer E., Maitl A., Rawat M., Youson J.H.,  
RA Keadley F.W.,  
RT "Characterization of lampyrin, an unusual matrix protein from lamprey  
RT cartilage. Implications for evolution, structure, and assembly of  
RT elastin and other fibrillar proteins."  
RL J. Biol. Chem. 268:1440-1447(1993).  
CC -1- FUNCTION: SELF-AGGREGATING PROTEIN THAT MAKES PART OF THE SOLUBLE  
CC FORM OF LAMPYRIN.  
CC -1- SUBUNIT: THE POLYMERIC LAMPYRIN CHAINS SELF-AGGREGATE TO FORM  
CC FIBERS AND HAVE SECONDARY STRUCTURES PARTICULARLY RICH IN BETA-  
CC SHEETS AND IN BETA-TURNS.

CC -1- SUBCELLULAR LOCATION: Extracellular matrix.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 0.9-12 (SHOWN HERE) AND 0.9-10;  
CC SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L05925: AAA49269.1;  
DR EMBL: L05924: AAA49268.1;  
DR PIR: C45051: C45051;  
DR PIR: A45051: A45051;  
KW Cartilage; Repeat; Connective tissue; Signal; Alternative splicing.  
FT SIGNAL 1 19  
FT CHAIN 20 139  
FT DOMAIN 42 110  
FT REPEAT 42 46  
FT REPEAT 47 51  
FT REPEAT 52 56  
FT REPEAT 57 61  
FT REPEAT 62 66  
FT REPEAT 67 71  
FT REPEAT 92 96  
FT REPEAT 106 110  
FT REPEAT 106 104  
FT VARSPPLIC 86 104  
FT SEQUENCE 139 AA: 13257 MW: E248AB7A6C6C7C3 CRC64;  
  
Query Match 7.2%; Score 8; DB 1; Length 139;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 ALLVALLL 13  
DB 7 ALLVALLL 14  
  
RESULT 3  
YXCC\_CYAPA STANDARD; PRT; 299 AA.  
ID YXCC\_CYAPA  
AC P31606;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 32.8 kDa protein in ycf13-apef intergenic region  
DE (ORF299).  
OS Cyanophora paradoxa.  
OC Cyanophyta.  
OC Glaucocystophyceae; Cyanophoraceae; Cyanophora.  
OX NCBI\_TaxID=2762;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LB555 / PRINGSHEIM;  
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohmert H.J.,  
RA Bryant D.A.,  
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa."  
RL Plant Mol. Biol. Rep. 13:327-332(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LB555 / PRINGSHEIM;  
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,  
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,  
RA Steiner J.M., Jakovitsch J., Bohmert H.J., Bryant D.A.,  
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:  
RT the genetic complexity of a primitive plastid."  
RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,  
RL Schweininger W. (eds.);  
RL Eukaryotism and Symbolism, pp.40-48, Springer-Verlag, Heidelberg  
RL (1997).

```

RN [3]
RP SEQUENCE OF 138-299 FROM N.A.
RX MEDLINE=92201692; PubMed=1551590;
RA Rhiet E., Stillewalt V.L., Gasparich G.E., Bryant D.A.;
RT "The psc genes of Synechococcus sp. PCC7002 and Cyanophora paradoxa:
  Cloning and sequence analysis."
RL Gene 112:123-128(1992).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE
  PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U30821; AAA81303.1; -.
DR PIR: M86239; AA65471.1; -.
DR PIR: PS0372; PS0372.
DR InterPro: IPR001626; ABC-3.
DR Pfam: PF00950; ABC-3; 1.
KM Hypothetical protein; Cyanelle; Transmembrane.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
SQ SEQUENCE 299 AA; 32817 MW; 7C8C30A20753CB66 CRC64;

Query Match
Best Local Similarity 7.2%; Score 8; DB 1; Length 299;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVIALLVY 15
DB 220 LVIALLVY 227

RESULT 4
Y4HA_RHISN STANDARD; PRT; 367 AA.
ID Y4HA_RHISN
AC P55471;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative ionic transporter Y4HA.
GN Y4HA.
OS Rhizobium sp. (strain NGR234).
OC Bacteri; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
  Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).
CC -1- FUNCTION: POSSIBLE CATION TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG. TO E.COLI CALCIUM/PROTON ANTIPORTER (CHAA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000075; AAB91689.1; -.
KM Hypothetical protein; Transport; Ion transport; Transmembrane;
  Plasmid.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
SQ SEQUENCE 367 AA; 38078 MW; 3AC0B829F2EFA2CB CRC64;

Query Match
Best Local Similarity 7.2%; Score 8; DB 1; Length 367;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVIAL 12
DB 224 GALLVIAL 231

RESULT 5
NAGA_VIRCH STANDARD; PRT; 378 AA.
ID NAGA_VIRCH
AC O32445; O9KTA9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) (GLCNAC 6-P
  deacetylase).
GN NAGA OR VC0994.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NON-O1 / 1148A;
RX MEDLINE=97446530; PubMed=9301118;
RA Yamano N., Oura N., Wang J., Fujishima S.;
RT "Cloning and sequencing of the genes for N-acetylglucosamine use that
  construct divergent operons (nagE-nagC) from Vibrio cholerae
  non-O1."
RL Biosci. Biotechnol. Biochem. 61:1349-1353(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Ginn M.L.,
  Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
  Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
  Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
  McDonald L., Uitterlinden T., Fleischmann R.D., Niernan W.C., White O.,
  Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
  Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
  cholerae."
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: N-acetyl-D-glucosamine 6-phosphate + H(2)O =
  D-glucosamine 6-phosphate + acetate.
CC -1- PATHWAY: N-ACERYL GLUCOSAMINE UTILIZATION PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE NAGA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```





DR EMBL: Y12380: CAA73023.1: JOINED.  
 DR EMBL: Y12383: CAA73023.1: JOINED.  
 DR EMBL: Y12384: CAA73023.1: JOINED.  
 DR EMBL: Y12385: CAA73023.1: JOINED.  
 DR EMBL: Y12386: CAA73023.1: JOINED.  
 DR EMBL: Y12387: CAA73023.1: JOINED.  
 DR EMBL: Y12388: CAA73023.1: JOINED.  
 DR EMBL: Y12389: CAA73023.1: JOINED.  
 DR EMBL: Y12390: CAA73023.1: JOINED.  
 DR EMBL: Y12382: CAA73023.1: JOINED.  
 DR EMBL: L23422: AAA16599.1: JOINED.  
 DR EMBL: Y12380: CAA73024.1: JOINED.  
 DR EMBL: Y12383: CAA73024.1: JOINED.  
 DR EMBL: Y12384: CAA73024.1: JOINED.  
 DR EMBL: Y12385: CAA73024.1: JOINED.  
 DR EMBL: Y12386: CAA73024.1: JOINED.  
 DR EMBL: Y12387: CAA73024.1: JOINED.  
 DR EMBL: Y12388: CAA73024.1: JOINED.  
 DR EMBL: Y12389: CAA73024.1: JOINED.  
 DR EMBL: Y12390: CAA73024.1: JOINED.  
 DR EMBL: Y12381: CAA73024.1: JOINED.  
 DR EMBL: Y12382: CAA73024.1: JOINED.  
 DR EMBL: L23421: AAA16598.1: JOINED.  
 DR HSSP: L16544: NOT\_ANNOTATED\_CDS.  
 DR HSSP: P11215: 1A8X.  
 DR MGD: MGI:102700:1:1ga7.  
 DR Interpro: IPR000413: Integrin\_alpha.  
 DR Pfam: PF01839: FG-GAP 5.  
 DR Pfam: PF00357: Integrin\_A 2.  
 DR PRINTS: PRO1185: INTEGRINA.  
 DR SMART: SM00191: Int\_alpha 5.  
 DR PROSITE: PS00242: INTEGRIN\_ALPHA 1.  
 DR Integrin: Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Repeat; Alternative splicing; ADP-ribosylation; Calcium  
 FT SIGNAL 1 33  
 FT CHAIN 34 1179 INTEGRIN ALPHA-7, HEAVY CHAIN (POTENTIAL).  
 FT CHAIN 34 955 INTEGRIN ALPHA-7, LIGHT CHAIN (POTENTIAL).  
 FT DOMAIN 34 1076 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1077 1102 POTENTIAL.  
 FT DOMAIN 1103 1179 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 50 87 FG-GAP 1.  
 FT REPEAT 122 154 FG-GAP 2.  
 FT REPEAT 194 225 FG-GAP 3.  
 FT REPEAT 305 341 FG-GAP 4.  
 FT REPEAT 362 400 FG-GAP 5.  
 FT REPEAT 423 453 FG-GAP 6.  
 FT REPEAT 481 521 FG-GAP 7.  
 FT CA\_BIND 372 380 POTENTIAL.  
 FT CA\_BIND 434 442 POTENTIAL.  
 FT CA\_BIND 492 500 POTENTIAL.  
 FT DOMAIN 953 958 POLY-ARG.  
 FT SITE 1105 1109 GEFER MOTIF.  
 FT DOMAIN 1155 1174 3 x 4 AA REPEATS OF D-X-H-P.  
 FT REPEAT 1155 1158 1.  
 FT REPEAT 1163 1166 2.  
 FT REPEAT 1171 1174 3.  
 FT DISULFID 94 103 BY SIMILARITY.  
 FT DISULFID 140 163 BY SIMILARITY.  
 FT DISULFID 184 197 BY SIMILARITY.  
 FT DISULFID 539 546 BY SIMILARITY.  
 FT DISULFID 552 615 BY SIMILARITY.  
 FT DISULFID 681 667 BY SIMILARITY.  
 FT DISULFID 781 792 BY SIMILARITY.  
 FT DISULFID 939 993 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 999 1004 BY SIMILARITY.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 784 784 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1043 1043 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 224 267 MISSING (IN ISOFORM ALPHA-7X2A AND ISOFORM ALPHA-7X2B).

FT VARSPLIC 268 307 MISSING (IN ISOFORM ALPHA-7X1A AND ISOFORM ALPHA-7X1B).  
 Query Match 7.28; Score 8; DB 1; Length 1179;  
 Best Local Similarity 100.0%; Pident. No. 6.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 LVYLVLLV 14  
 DB 1091 LVYLVLLV 1098  
 RESULT 8  
 ID ITA7\_HUMAN STANDARD: PRT; 1181 AA.  
 AC Q13683; Q43197; Q9UEV2; Q9UEF0; Q9N899;  
 DT 16-OCT-2001 (rel. 40, Created)  
 DT 16-OCT-2001 (rel. 40, Last sequence update)  
 DT 01-MAR-2002 (rel. 41, Last annotation update)  
 DE Integrin alpha-7 precursor.  
 GN ITGA7  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS ALPHA-7X2B AND ALPHA-7X2DB).  
 RC TISSUE-Petal heart, and Osteoblast;  
 RA MEDLINE-9813911; PubMed-9473524;  
 RS Leung E., Lim S.P., Berg R., Yang Y., Ni J., Wang S.-X.,  
 RA Kristiansen G.W.;  
 RT "A novel extracellular domain variant of the human integrin alpha 7 subunit generated by alternative intron splicing";  
 RL Biochem. Biophys. Res. Commun. 243:317-325(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2B).  
 RC MEDLINE-98250181; PubMed-9590289;  
 RA Hirasawa Y.K., Chou F.-L., Engvall E., Ogawa M., Matsuda C.,  
 RS Hirabayashi S., Yokochi K., Zlobet B.L., Kramer R.H., Kaufman S.J.,  
 RA Orsane E., Goto Y.-I., Nonaka I., Tsukihara T., Wang D.Z.;  
 RA Hoffman B.P., Arachata K.;  
 RT "Mutations in the integrin alpha7 gene cause congenital myopathy";  
 RL Nat. Genet. 19:94-97(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2B).  
 RC Viziathanakis I.S., Zlobet B.L., Kramer R.H.;  
 RA "Cloning of human integrin alpha-7 cDNA";  
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RC TISSUE-Skeletal muscle;  
 RA MEDLINE-99333684; PubMed-10403775;  
 RS Vignier N., Moghadaszadeh B., Gary F., Beckmann J., Mayer U.,  
 RA Guicheney P.;  
 RT "Structure, genetic localization, and identification of the cardiac and skeletal muscle transcripts of the human integrin alpha7 gene (ITGA7)";  
 RL Biochem. Biophys. Res. Commun. 260:357-364(1999).  
 RN [5]  
 RP SEQUENCE OF 34-45.  
 RC TISSUE-Melanoma;  
 RA MEDLINE-92198982; PubMed-1839357;  
 RS Kramer R.H., Vu M.P., Cheng Y.F., Ramos D.M., Timpl R., Waleh N.;  
 RA "Laminin-binding Integrin alpha 7 beta 1: functional characterization and expression in normal and malignant melanocytes";  
 RL Cell Regul. 2:805-817(1991).  
 RN [6]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM ALPHA-7X2).  
 RC TISSUE-Heart;  
 RA MEDLINE-94075378; PubMed-8253814;  
 RS Zlobet B.L., Vu M.P., Waleh N., Crawford J., Lin C.-S., Kramer R.H.;  
 RA "Alternative extracellular and cytoplasmic domains of the integrin alpha 7 subunit are differentially expressed during development";

RL J. Biol. Chem. 268:26773-26783(1993).  
RN [7]  
RP SEQUENCE OF 1105-1181 FROM N.A. (ISOFORM ALPHA-7B).  
RC TISSUE-Fetal muscle;  
RX MEDLINE-94171924; PubMed-8126096;  
RA Song W.K., Wang W., Sato H., Bielser D.A., Kaufman S.J.;  
RT "Expression of alpha 7 integrin cytoplasmic domains during skeletal  
muscle development: alternate forms, conformational change, and  
homologues with serine/threonine kinases and tyrosine phosphatases.";  
RL J. Cell Sci. 106:1139-1152(1993).  
RN [8]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM ALPHA-7A).  
RC TISSUE-Skeletal muscle;  
RX MEDLINE-98012902; PubMed-9352853;  
RA Basore N., Vachon P.H., Herring-Gillam F.E., Perreault N.,  
Beaulieu J.-F.;  
RT "Relation between integrin alpha7/beta1 expression in human intestinal  
cells and enterocytic differentiation.";  
RL Gastroenterology 113:1510-1521(1997).  
RN [9]  
RP FUNCTION.  
RX MEDLINE-97453329; PubMed-9307969;  
RA Ziober B.L., Chen Y.O., Kramer R.H.;  
RT "The laminin-binding activity of the alpha 7 integrin receptor is  
defined by developmentally regulated splicing in the extracellular  
domain.";  
RL Mol. Biol. Cell 8:1723-1734(1997).  
RN [10]  
RP FUNCTION.  
RX MEDLINE-20160722; PubMed-10694445;  
RA Schoeber S., Melenz D., Echtermeier F., Hapke S., Poeschl E.,  
von der Mark H., Moch H., von der Mark K.;  
RT "The role of extracellular and cytoplasmic splice domains of  
alpha7-integrin in cell adhesion and migration on laminins.";  
RL Exp. Cell Res. 235:303-313(2000).  
RN [11]  
RP TISSUE SPECIFICITY.  
RX MEDLINE-96197133; PubMed-8626012;  
RA Martin P.T., Kaufman S.J., Kramer R.H., Sanes J.R.;  
RT "Synaptic integrins in developing, adult, and mutant muscle: selective  
association of alpha1, alpha7A, and alpha7B integrins with the  
neuromuscular junction.";  
RL Dev. Biol. 174:125-139(1996).  
RN [12]  
RP FUNCTION.  
RX MEDLINE-97101966; PubMed-97101966;  
RA "FUNCTION: INTEGRIN ALPHA-7/BETA-1 IS THE PRIMARY LAMININ RECEPTOR  
ON SKELETAL MYOBLASTS AND ADULT MYOFIBERS. DURING MYOGENIC  
DIFFERENTIATION, IT MAY INDUCE CHANGES IN THE SHAPE AND MOBILITY  
OF MYOBLASTS, AND FACILITATE THEIR LOCALIZATION AT LAMININ-RICH  
SITES OF SECONDARY FIBER FORMATION. IT IS INVOLVED IN THE  
MAINTENANCE OF THE MYOFIBERS CYTOARCHITECTURE AS WELL AS FOR THEIR  
ANCHORAGE, VIABILITY AND FUNCTIONAL INTEGRITY. ISOFORMS ALPHA-7X2B  
AND ALPHA-7X1B PROMOTE MYOBLAST MIGRATION ON LAMININ 1 AND LAMININ  
2/4, BUT ALPHA-7X1B IS LESS ACTIVE ON LAMININ 1 (IN VITRO).  
SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA  
SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A  
DISULFIDE BOND. ALPHA-7 ASSOCIATES WITH BETA-1.  
SUBCELLULAR LOCATION: Type 1 membrane protein.  
ALTERNATIVE PRODUCTS: AT LEAST 12 ISOFORMS: ALPHA-7X1A, ALPHA-  
7X1B, ALPHA-7X1D, ALPHA-7X1E, ALPHA-7X2A, ALPHA-7X2B, ALPHA-  
7X2C, ALPHA-7X2D, ALPHA-7X1X2A, ALPHA-7X1X2B (SHOWN HERE),  
ALPHA-7X1X2D, AND ALPHA-7X1X2E; MAY BE PRODUCED BY ALTERNATIVE  
SPLICING. THERE IS A COMBINATION OF AT LEAST FIVE ALTERNATIVE  
CYTOPLASMIC DOMAINS, THREE EXTRACELLULAR (X1, X2 AND D) AND TWO  
CYTOPLASMIC (A AND B). A THIRD POTENTIAL ALTERNATIVE SPLICED  
CYTOPLASMIC DOMAIN (C) DOES NOT APPEAR TO BE EXPRESSED. IN VITRO  
GENERATED ISOFORM X2C SHOWS FUNCTION. SO FAR DETECTED ARE ISOFORMS  
ALPHA-7X1A, ALPHA-7X2B (MAJOR), ALPHA-7X2D (MINOR) AND MINOR  
ISOFORMS CONTAINING SEGMENT XIX2.  
TISSUE SPECIFICITY: ISOFORMS CONTAINING SEGMENT A ARE  
PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLE. ISOFORMS CONTAINING  
SEGMENT B ARE ABUNDANTLY EXPRESSED IN SKELETAL MUSCLE, MODERATELY  
IN CARDIAC MUSCLE, SMALL INTESTINE, COLON, OVARY AND PROSTATE AND  
WEALY IN LUNG AND TESTES. ISOFORMS CONTAINING SEGMENT X2D ARE  
EXPRESSED AT LOW LEVELS IN FETAL AND ADULT SKELETAL MUSCLE AND IN

CC CARDIAC MUSCLE, BUT ARE NOT DETECTED IN MYOBLASTS AND MYOTUBES. IN  
CC MUSCLE FIBERS ISOFORMS CONTAINING SEGMENT A AND B ARE EXPRESSED AT  
CC MYOTENDINOSUS AND NEUROMUSCULAR JUNCTIONS. ISOFORMS CONTAINING  
CC SEGMENT C ARE EXPRESSED AT NEUROMUSCULAR JUNCTIONS AND AT  
CC EXTRASYNAPTIC SITES. ISOFORMS CONTAINING SEGMENTS X1 OR X2 OR, AT  
CC LOW LEVELS, XIX2 ARE EXPRESSED IN FETAL AND ADULT SKELETAL MUSCLE  
CC (MYOBLASTS AND MYOTUBES) AND CARDIAC MUSCLE.  
CC DEVELOPMENTAL STAGE: IN RENOVATING INTESTINAL EPITHELIUM, EXPRESSION  
CC OF ISOFORMS CONTAINING SEGMENT B CORRELATES WITH THE ONSET OF  
CC ENTEROCYTIC DIFFERENTIATION.  
CC PTM: ADP-RIBOSYLATED ON AT LEAST TWO SITES OF THE EXTRACELLULAR  
CC DOMAIN IN SKELETAL MYOTUBES (BY SIMILARITY).  
CC DISEASE: DEFECTS IN ITGA7 ARE ASSOCIATED WITH A FORM OF CONGENITAL  
CC MYOPATHY: A GROUP OF HETEROGENEOUS MUSCLE DISORDERS WHICH ARE  
CC THOUGHT TO RESULT FROM ABNORMAL MUSCLE DEVELOPMENT. MUSCLE  
CC WEAKNESS IS EITHER NON-PROGRESSIVE OR SLOWLY PROGRESSIVE AND  
CC APPARENT FROM BIRTH OR EARLY INFANCY.  
CC SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL; AF032108; AAC39708.1; -;  
CC EMBL; AF052050; AAC18968.1; -;  
CC EMBL; AF072132; AAC80458.1; -;  
CC EMBL; AJ228836; CAB41534.1; -;  
CC EMBL; AJ228837; CAB41534.1; JOINED.  
CC EMBL; AJ228838; CAB41534.1; JOINED.  
CC EMBL; AJ228839; CAB41534.1; JOINED.  
CC EMBL; AJ228840; CAB41534.1; JOINED.  
CC EMBL; AJ228842; CAB41534.1; JOINED.  
CC EMBL; AJ228843; CAB41534.1; JOINED.  
CC EMBL; AJ228844; CAB41534.1; JOINED.  
CC EMBL; AJ228845; CAB41534.1; JOINED.  
CC EMBL; AJ228846; CAB41534.1; JOINED.  
CC EMBL; AJ228847; CAB41534.1; JOINED.  
CC EMBL; AJ228848; CAB41534.1; JOINED.  
CC EMBL; AJ228849; CAB41534.1; JOINED.  
CC EMBL; AJ228850; CAB41534.1; JOINED.  
CC EMBL; AJ228851; CAB41534.1; JOINED.  
CC EMBL; AJ228852; CAB41534.1; JOINED.  
CC EMBL; AJ228853; CAB41534.1; JOINED.  
CC EMBL; AJ228854; CAB41534.1; JOINED.  
CC EMBL; AJ228855; CAB41534.1; JOINED.  
CC EMBL; AJ228856; CAB41534.1; JOINED.  
CC EMBL; AJ228857; CAB41534.1; JOINED.  
CC EMBL; AJ228858; CAB41534.1; JOINED.  
CC EMBL; AJ228859; CAB41534.1; JOINED.  
CC EMBL; AJ228860; CAB41534.1; JOINED.  
CC EMBL; AJ228862; CAB41534.1; JOINED.  
CC EMBL; AJ228863; CAB41535.1; JOINED.  
CC EMBL; AJ228837; CAB41535.1; JOINED.  
CC EMBL; AJ228838; CAB41535.1; JOINED.  
CC EMBL; AJ228839; CAB41535.1; JOINED.  
CC EMBL; AJ228841; CAB41535.1; JOINED.  
CC EMBL; AJ228842; CAB41535.1; JOINED.  
CC EMBL; AJ228843; CAB41535.1; JOINED.  
CC EMBL; AJ228844; CAB41535.1; JOINED.  
CC EMBL; AJ228845; CAB41535.1; JOINED.  
CC EMBL; AJ228846; CAB41535.1; JOINED.  
CC EMBL; AJ228847; CAB41535.1; JOINED.  
CC EMBL; AJ228848; CAB41535.1; JOINED.  
Query Match 7.2%; Score 8; DB 1; Length 1181;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 7 LVLVALL 14
DB 1093 LVLVALL 1100

RESULT 9
YEDD_SALTY STANDARD: PRT; 141 AA.
ID YEDD_SALTY STANDARD: PRT; 141 AA.
AC Q06399;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Hypothetical lipoprotein yedd precursor.
OS Salmoella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SW1103;
RX MEDLINE-93381452; PubMed-8371104;
RA Rana M., Kihara M., Kawagishi T., Macnab R.M.;
RT "Organization of the Escherichia coli and Salmoneilla typhimurium
chromosomes between flagellar regions iira and iirb, including a
large non-coding region."
RL J. Gen. Microbiol. 139:1401-1407 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-1167609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
LT2."
RL Nature 413:852-856 (2001).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L13280; AAA71971.1;
CC DR EMBL: A608787; AAL20876.1;
CC DR STrGene: SC10420; yedd.
CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN_1.
CC KW Hypothetical protein; Membrane; Lipoprotein; Signal;
CC Complete proteome.
CC FT SIGNAL 1 15
CC FT CHAIN 16 141
CC FT LIPID 16 16
CC FT CONFLICT 123 124 CA->WP (IN REF. 1).
CC FT SEQUENCE 141 AA; 15356 MW; A362EBBBD9C8DD9 CRC64;
CC SQ

Query Match 6.3%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11
DB 8 GALLVLA 14

RESULT 10

LVC2_PIG
ID LVC2_PIG STANDARD: PRT; 146 AA.
AC P12068;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Lysozyme C-2 precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase C).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96292525; PubMed-8728388;
RA Yu M., Irwin D.M.;
RT "Evolution of stomach lysozyme: the pig lysozyme gene."
RL Mol. Phylogenet. Evol. 5:298-308 (1996).
RN [2]
RP SEQUENCE OF 19-146.
RC TISSUE=Stomach;
RX MEDLINE-89362490; PubMed-2504928;
RA Jolles J., Jolles P., Bowman B.H., Prager E.M., Stewart C.-B.,
Wilson A.C.;
RT "Episodic evolution in the stomach lysozymes of ruminants."
RL J. Mol. Evol. 28:528-535 (1989).
CC -1- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION. THOSE
IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOGENS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
heteropolymers of the prokaryotes cell walls.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: LYSOZYMES C ARE CAPABLE OF BOTH HYDROLYSIS &
TRANSEGLYCOSYLATION; THEY SHOW ALSO A SLIGHT ESTERASE ACTIVITY.
THEY ACT RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED & UNSUBSTITUTED
PEPTIDOLYCAN 6, SLOWLY, ON CHITIN OLIGOSACCHARIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U28757; AAB16862.1;
CC DR HSSP: P00695; 3LHM.
CC DR InterPro: IPR001916; Lactabmn_lysozyme.
CC DR Pfam: PF00062; Lys_1.
CC DR PRINTS: PR00135; LYZLACT.
CC DR SMART: SM00263; LYZ1.1.
CC DR PROSITE: PS00128; LACTALBUMIN_LYSOZYME_1.
CC KW Hydrolyase; Glycosidase; Bacteriolytic enzyme; Digestion; Stomach;
CC Signal.
CC FT SIGNAL 1 18
CC FT CHAIN 19 146
CC FT DISULFID 24 144
CC FT DISULFID 48 132
CC FT DISULFID 81 97
CC FT DISULFID 93 111
CC FT ACT_SITE 53 53
CC FT ACT_SITE 69 69
CC FT SEQUENCE 146 AA; 16484 MW; C2986F5CAAF9F4 CRC64;
CC SQ

Query Match 6.3%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLVALL 13
DB 4 LVLVALL 10

```



RESULT 11  
LVC3\_PIG STANDARD; PRT; 148 AA.  
AC P12069;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Lysozyme C-3 precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase C).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
ON NCBI\_TaxID=9823;  
RX MEDLINE=9362490; PubMed=2504928;  
RA Jolles J., Jolles P., Bowman B.H., Prager E.M., Stewart C.-B.,  
RA Wilson A.C.;  
RT "Episodic evolution in the stomach lysozymes of ruminants.";  
RL J. Mol. Evol. 28:528-535(1989).  
RP SEQUENCE OF 19-148.  
RC TISSUE=Stomach;  
RN Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC  
CC -1- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION. THOSE  
CC IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-  
CC MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-  
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan  
CC heteropolymers of the prokaryotes cell walls.  
CC -1- SUBUNIT: MONOMER.  
CC -1- MISCELLANEOUS: LYSOZYMES C ARE CAPABLE OF BOTH HYDROLYSIS &  
CC TRANSGLYCOSYLATION; THEY SHOW ALSO A SLIGHT ESTERASE ACTIVITY;  
CC THEY ACT RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED & UNSUBSTITUTED  
CC PEPTIDOLYCAN &, SLOWLY, ON CHITIN OLIGOSACCHARIDES.  
CC -1- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U44435; AA06644.1; -.  
DR PIR: S10047; S10047.  
DR HSSP: P00695; 3IHM.  
DR InterPro: IPR001916; Lactaldmn\_Lysozyme.  
DR Pfam: PF00062; Lys\_1.  
DR PRINTS: PR00135; LYZACT.  
DR SMART: SM00263; LYZ1.1.  
DR PROSITE: PS00128; LACTALBUMIN\_LYSOZYME; 1.  
KM Hydrolyase; Glycosidase; Bacteriolytic enzyme; Digestion; Stomach;  
KW SIGNAL.  
FT CHAIN 1 18  
FT DISULFID 19 148  
FT DISULFID 24 146  
FT DISULFID 48 134  
FT DISULFID 83 99  
FT DISULFID 95 113  
FT ACT\_SITE 53 53  
FT ACT\_SITE 71 71  
SQ SEQUENCE 148 AA; 16711 MW; 76AA67BFABBD64E6 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 LVLVALL 13

Db 4 LVLVALL 10  
RESULT 12  
BACR\_HALS4 STANDARD; PRT; 250 AA.  
ID BACR\_HALS4  
AC O93740;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Bacteriorhodopsin (BR).  
GN BOP.  
OS Halobacterium sp. (strain arg-4).  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
ON NCBI\_TaxID=160432;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99096913; PubMed=9878396;  
RA Ihara K., Umemura T., Kataagiri I., Kitaajima-Ihara T., Sugiyama Y.,  
RA Kimura Y., Mukohata Y.;  
RT "Evolution of the archaeal rhodopsins: evolution rate changes by gene  
RT duplication and functional differentiation.";  
RL J. Mol. Biol. 285:163-174(1999).  
CC  
CC -1- FUNCTION: BACTERIORHODOPSIN IS A LIGHT-DRIVEN PROTON PUMP.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE ARCHAEL OPSIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AB009620; BAA75200.1; -.  
DR HSSP: P02945; 1F50.  
DR InterPro: IPR001425; Bac\_Rhodopsin.  
DR Pfam: PF01036; Bac\_Rhodopsin; 1.  
DR PRINTS: PR00251; BACTRHODOPSIN.  
DR PROSITE: PS00950; BACTERIAL\_OPSIN\_1; 1.  
DR PROSITE: PS00327; BACTERIAL\_OPSIN\_RET; 1.  
KM Ion transport; Photoreceptor; Transmembrane; Retinal protein;  
KW Hydrogen ion transport.  
FT DOMAIN 1 16  
FT TRANSMEM 17 37  
FT TRANSMEM 52 72  
FT TRANSMEM 91 111  
FT TRANSMEM 115 135  
FT TRANSMEM 143 163  
FT TRANSMEM 185 205  
FT TRANSMEM 213 233  
FT DOMAIN 234 250  
FT BINDING 225 225  
SQ SEQUENCE 250 AA; 27041 MW; B72D3373506FD275 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 250;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 42 LLDLDSL 48  
Db 101 LLDLDSL 107  
RESULT 13  
QOX2\_BACSU STANDARD; PRT; 322 AA.  
ID QOX2\_BACSU  
AC P34957; O32281;  
DT 01-FEB-1994 (Rel. 28, Created)

30-MAY-2000 (Rel. 39, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 Quinol oxidase polypeptide II precursor (EC 1.9.3.-) (Quinol oxidase  
 Aa3-600, subunit qoxA) (Oxidase Aa(3) subunit 2).  
 GN QOXA OR IPX-37D.  
 OC Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OK NCBI\_TaxID=1423;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA MEDLINE=92268053; PubMed=1316894;  
 RA Santana M., Kunst F., Hullo M.F., Rapoport G., Danchin A.,  
 RA Glaser P.,  
 RT Molecular cloning, sequencing, and physiological characterization of  
 RT the qox operon from Bacillus subtilis encoding the aas-600 quinol  
 RT oxidase.  
 RT J Biol. Chem. 267:10225-10231(1992).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA MEDLINE=95020537; PubMed=7934828;  
 RA Glaser P., Kunst F., Arnaut M., Goudart M.P., Gonzales W.,  
 RA Hullo M.F., Ionescu W., Lubchinsky B., Marcelino L., Moszer I.,  
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
 RA Rapoport G., Danchin A.,  
 RT Bacillus subtilis genome project: cloning and sequencing of the 97  
 RT kb region from 325 degrees to 333 degrees.  
 RT M. Microbiol. 10:371-384(1993).  
 RL 13  
 RS REFERENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,  
 RA Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.  
 RL 4)  
 RS REFERENCE OF 27-43, AND CHARACTERIZATION.  
 RC STRAIN=423;  
 RA MEDLINE=9609139; PubMed=757098;  
 RA Lemma E., Simon J., Schagger H., Krogger A.,  
 RT Properties of the menaquinol oxidase (qox) and of qox deletion  
 RT mutants of Bacillus subtilis.  
 RL 1)  
 RS REFERENCE OF THE MENAQUINOL OXIDATION WITH THE CONCOMITANT  
 CC REDUCTION OF OXYGEN TO WATER. MAJOR COMPONENT FOR ENERGY  
 CC CONVERSION DURING VEGETATIVE GROWTH. SUBUNIT II TRANSFERS THE  
 CC ELECTRONS FROM A QUINOL TO THE NUCLEAR CENTER OF THE CATALYTIC  
 CC SUBUNIT I.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: SOME TO MITOCHONDRIAL OR BACTERIAL COX2 SUBUNITS:  
 CC BUT LACK HEME-BINDING DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M86548; AAA22686.1; ALT\_INIT.  
 CC EMBL: X73124; CAAS1593.1; ALT\_INIT.  
 CC EMBL: Z89123; CAB15843.1; ALT\_INIT.  
 CC PIR: A38129; A38129.  
 CC HSP: P18400; IFTT.  
 CC Subtilist; BG10583; qoxA.  
 CC Interpro: IPR001505; COX2.  
 CC Interpro: IPR002425; Cyl\_cox\_2.  
 CC Pfam: PF00116; COX2\_1.  
 CC Pfam: PF02790; COX2\_TM\_1.  
 CC PRINTS: PR01166; CYCOXIDASE1.  
 CC PRODOM: PD000131; COX2\_1.  
 CC PROSITE: PS00013; PROXAR\_LIPOPROTEIN\_1.

Query Match 6.3%; Score 7; DB 1; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 LVLALL 13  
 DB 12 LVLALL 18  
 RESULT 14  
 CIB2\_CAVPO STANDARD; PRT; 332 AA.  
 ID 090221;  
 AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE T-cell surface glycoprotein CD1b2 precursor (CD1-b2 antigen).  
 GN CD1B2.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavidae; Cavia.  
 ON NCBI\_TaxID=10141;  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HARTLEY, AND NIH 2; TISSUE=Thymus;  
 RX MEDLINE=20021845; PubMed=10553074;  
 RA Dascher G.C., Hirumatsu K., Naylor J.W., Brauer P.P., Brown K.A.,  
 RA Stclair K.P., Behar S.M., Kawasaki E.S., Porcelli S.A., Brenner M.B.,  
 RT Conservation of a CD1 multigene family in the guinea pig.  
 RL J. Immunol. 163:5478-5488(1999).  
 CC -1- FUNCTION: NOT KNOWN.  
 CC -1- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF145484; AAF12739.1;  
 CC Interpro: IPR003006; IG\_MHC.  
 CC Interpro: IPR003597; IG-cl.  
 CC Pfam: PF00047; Ig1.  
 CC SMART: SM00407; IGc1.1.  
 CC Glycoprotein; Signal; Transmembrane; Immunoglobulin domain;  
 CC Multigene family.  
 CC SIGNAL 1 17  
 CC CHAIN 18 332  
 CC DOMAIN 18 301  
 CC TRANSMEM 302 332  
 CC DOMAIN 323 332  
 CC DISULFID 120 184  
 CC DISULFID 224 279  
 CC CARBOHYD 38 38  
 CC CARBOHYD 75 75  
 CC CARBOHYD 146 146  
 CC  
 CC POTENTIAL.  
 CC T-CELL SURFACE GLYCOPROTEIN CD1B2.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).  
 CC BY SIMILARITY.  
 CC BY SIMILARITY.  
 CC N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC N-LINKED (GLCNAC... ) (POTENTIAL).  
 CC N-LINKED (GLCNAC... ) (POTENTIAL).



CC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=11773;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98295987; PubMed=9634230;  
 RX STRAIN=H37RV;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Horsby T., Jagsen K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fietischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouli H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC -1- CATALYTIC ACTIVITY: N2-acetyl-L-ornithine + 2-oxoglutarate = N-  
 CC acetyl-L-glutamate 5-semialdehyde + L-glutamate.  
 CC  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC  
 CC -1- PATHWAY: FOURTH STEP IN ARGININE BIOSYNTHESIS.  
 CC  
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: Z85982; CAB06649.1;  
 CC EMBL: AE007032; AAK5962.1;  
 CC HSSP: P04181; ZOAT.  
 CC TIGR: MT1693;  
 DR TubercuList; RV1655;  
 DR InterPro: IPR000954; AminoTran\_3.  
 DR Pfam: PF00202; AminoTran\_3; 1.  
 DR PROSITE: PS00600; AA\_TRANSFERR\_CLASS\_3; 1.  
 DR Transferrase; AminoTransferase; Pyridoxal phosphate;  
 KW Arginine biosynthesis; Complete proteome.  
 FT BINDING 253  
 FT PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 400 AA; 40910 MW; 833846D52979019 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 400;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 RGALLVL 10  
 DB 217 RGALLVL 223

RESULT 18  
 HISZ\_SYNP7 STANDARD; PRT; 420 AA.  
 AC 055267;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ATP phosphoribosyltransferase regulatory subunit.  
 GN HISZ.  
 OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1140;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95050240; PubMed=7961432;  
 RX Tsinoremas N.F., Kutach A.K., Strayer C.A., Golden S.S.;  
 RA "Efficient gene transfer in Synechococcus sp. strains PCC 7942 and  
 RT PCC 6301 by interspecies conjugation and chromosomal recombination.";  
 RL J. Bacteriol. 176:6764-6768(1994).  
 CC  
 CC -1- FUNCTION: MAY ALLOW THE REGULATION OF ATP  
 CC PHOSPHORIBOSYLTRANSFERASE ACTIVITY BY HISTIDINE (BY SIMILARITY).  
 CC  
 CC -1- PATHWAY: FIRST STEP IN HISTIDINE BIOSYNTHETIC PATHWAY. IS VERY  
 CC IMPORTANT IN THE REGULATION OF HISTIDINE METABOLISM.  
 CC  
 CC -1- SUBUNIT: BINDS TO HISG (BY SIMILARITY).  
 CC  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC  
 CC -1- MISCELLANEOUS: THIS FUNCTION IS GENERALLY FULFILLED BY THE C-  
 CC TERMINAL PART OF HISG, WHICH IS MISSING IN SOME BACTERIA SUCH AS  
 CC THIS ONE.  
 CC  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC STRONG, TO HISTIDYL-TRNA SYNTHETASES.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: U35476; AAA64445.1; AA-TRNA\_Ligase\_II.  
 CC InterPro: IPR002106; AA-TRNA\_Ligase\_II.  
 CC DR InterPro: IPR002314; tRNA\_Synt\_2D.  
 CC DR Pfam: PF00587; tRNA\_Synt\_2D.  
 CC DR PROSITE: PS00178; AA-TRNA\_Ligase\_II; 1. FALSE NEG.  
 CC DR PROSITE: PS00333; AA-TRNA\_Ligase\_II.2; FALSE NEG.  
 KW Histidine biosynthesis456  
 SQ SEQUENCE 420 AA; 46568 MW; C4938788776961 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 LILDLST 48  
 DB 261 LILDLST 267

RESULT 19  
 ORB2\_BACSU STANDARD; PRT; 424 AA.  
 AC P37942;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Liponamide acyltransferase component of branched-chain alpha-keto acid  
 DE dehydrogenase complex (EC 2.3.1.-) (E2) (Dihydroliponamide branched  
 DE chain transferase).  
 GN BMBB OR BMB2 OR BMB.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/clostridium group;  
 CC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=93279308; PubMed=8504804; Kanada T.;  
 RA Wang G.-F., Kuriki T., Roy K.L., Kanada T.;  
 RT "The primary structure of branched-chain alpha-oxo acid dehydrogenase  
 RT from Bacillus subtilis and its similarity to other alpha-oxo acid

RT dehydrogenases.";  
 RL Eur. J. Biochem. 213:1091-1099(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,  
 RL Sato T., Takeuchi M.;  
 RN Submitted (May-1996) to the EMBL/Genbank/DBJ databases.  
 [3]  
 RP SEQUENCE OF 266-424 FROM N.A.  
 RC STRAIN-168 / MARBURG;  
 RA MEDLINE-95050642; PubMed-7961792;  
 RX Ahmed M., Borsch C.M., Taylor S.S., Vazquez-Laslop N., Neyfakh A.A.;  
 RT "A protein that activates expression of a multidrug efflux  
 transporter upon binding the transporter substrates.";  
 RL J. Biol. Chem. 269:28506-28513(1994).  
 CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX  
 CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA  
 AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS:  
 CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE  
 CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).  
 CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL  
 CC COFACTOR.  
 CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL  
 CC SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M97391; AAA2280.1; -  
 DR EMBL: D84432; BAA12600.1; -  
 DR EMBL: L25604; AAB81541.1; -  
 DR EMBL: 299116; CAB14334.1; -  
 DR PIR: S32488; S32488.  
 DR HSSP: P07016; 1E2O.  
 DR Subtilisin; BG10305; bfmB.  
 DR InterPro: IPR001078; 2oxoacid\_dh.  
 DR InterPro: IPR000089; Biotin\_lipoyl.  
 DR InterPro: IPR003016; lipoyl  
 DR InterPro: IPR004167; e3\_binding.  
 DR Pfam: PF00198; 2-oxoacid\_dh; 1.  
 DR Pfam: PF00364; biotin\_lipoyl; 1.  
 DR Pfam: PF02817; e3\_binding; 1.  
 DR Prodom: PD001115; 2oxoacid\_dh; 1.  
 DR PROSITE: PS00189; LIPOYL\_1.  
 DR Glycolysis; Transferase; Acyltransferase; Lipoyl; Complete proteome.  
 FT BINDING 44 44 LIPOYL (POTENTIAL).  
 FT ACT\_SITE 395 395 POTENTIAL.  
 FT ACT\_SITE 399 399 POTENTIAL.  
 FT CONFLICT 305 305 A -> P (IN REF. 3).  
 FT CONFLICT 350 350 V -> I (IN REF. 3).  
 FT CONFLICT 365 365 Q -> E (IN REF. 3).  
 FT CONFLICT 370 370 V -> D (IN REF. 3).  
 SQ SEQUENCE 424 AA; 45837 MW; 2E0C9EAF817CE228 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 424;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 76 RVLDDGLV 82  
 DB 396 RVLDDGLV 402  
 RESULT 20

GSA\_CERSY  
 ID GSA\_CERSY STANDARD; PRT; 434 AA.  
 AC 074038;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)  
 DE (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).  
 GN HEML OR GSAF.  
 OS Cenarchaeum symbiosum.  
 OC Archaea; Cenarchaeota; Cenarchaeum.  
 OX NCBI\_TaxID=46770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A;  
 RX MEDLINE-98422450; PubMed-9748430;  
 RA Schlieper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,  
 RA Swanson R.V.;  
 RT "Genomic analysis reveals chromosomal variation in natural populations  
 of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";  
 RL J. Bacteriol. 180:5003-5009(1998).  
 CC -1- CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate -> 5-  
 CC aminolevulinate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY; SECOND STEP.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF083071; AAC62681.1; -  
 DR HSSP: P24630; 2GSA.  
 DR InterPro: IPR000954; Aminotran\_3.  
 DR Pfam: PF00202; aminotran\_3; 2.  
 DR PROSITE: PS00600; AA\_TRANSFER\_CLASS\_3; FALSE\_NEG.  
 DR Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate.  
 FT BINDING 267 267 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 434 AA; 46714 MW; D1ED70D39A25CA50 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 434;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 RGALLVL 10  
 DB 232 RGALLVL 238  
 RESULT 21  
 ID RG60\_DEIRA STANDARD; PRT; 531 AA.  
 AC 09R0W8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 60-kDa SS-A/Ro ribonucleoprotein homolog (Ro sixty related).  
 GN RSR OR DR1262.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RL;  
 RX MEDLINE-20036896; PubMed-10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,



FT CONFLICT 437 437 L -> F (IN REF. 3 AND 4).  
 FT CONFLICT 438 P -> A (IN REF. 4).  
 FT CONFLICT 507 IAE -> A (IN REF. 1 AND 3).  
 SO SEQUENCE 574 AA; 62984 MW; B62DEF64338CB8C CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 1; Length 574;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ALGVKMA 23  
 |||||  
 Db 430 ALGVKMA 436

RESULT 23  
 ID ILVI\_SALTY STANDARD; PRT; 574 AA.

AC P40811;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Acetolactate synthase I (large subunit) (EC 4.1.3.18) (AHAS-III) (Acetohydroxy-acid synthase III large subunit) (ALS-III).  
 GN ILVI OR STM0116.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxId=602;

RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2 / SGGC1412 / ATCC 700720;  
 RX MEDLINE-21534948; PubMed-11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.";  
 RT Nature 413:852-856(2001).

RL Nature 413:852-856(2001).

RN SEQUENCE OF 516-574 FROM N.A.

RP STRAIN-LT2;

RX MEDLINE-91238714; PubMed-1851954;

RA Jandreis K., Postma P.W., Lengeler J.W.;  
 RT "Nucleotide sequence of the *ilvI*-*trp* gene region of *Escherichia coli* K12 and *Salmonella typhimurium* LT2.";  
 RL Mol. Gen. Genet. 226:332-336(1991).

CC -1- CATALYTIC ACTIVITY: 2-ACETOACETATE + CO(2) = 2 PYRUVATE (THIS ENZYME ALSO CATALYZES FORMATION OF 2-ACETO-2-HYDROXYBUTANOATE).

CC -1- COFACTOR: THIAMINE PYROPHOSPHATE, AND MAGNESIUM.

CC -1- ENZYME REGULATION: SENSITIVE TO VALINE INHIBITION.

CC -1- PATHWAY: FIRST STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS.

CC -1- SUBUNIT: DIMER OF CHAIN H AND CHAIN I.

CC -1- MISCELLANEOUS: S. TYPHIMURIOUM CONTAINS GENES FOR 3 AHAS ISOZYMES: ILVBN, ILVGM AND ILVIH.

CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.

CC -1- CAUTION: A stop codon in position 12 was translated as Arg to restore the similarity with the N-terminal region of others *ilvi* homologs.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AE008699; AAL19080.1; ALT\_INIT.

DR EMBL: X55456; CAA39101.1; -  
 DR PIR: S15939; S15939.

DR StyGene; SG10129; ILVI.  
 DR InterPro; IPR000399; TPP-enzyme.  
 DR PROSITE; PS00187; TPP-ENZYMES; 1.  
 KW Branched-chain amino acid biosynthesis; Flavoprotein; Magnesium;  
 KW Thiamine pyrophosphate; Lyase; Complete proteome.  
 FT ACT\_SITE 51  
 FT ACT\_SITE 51 BY SIMILARITY.  
 SO SEQUENCE 574 AA; 62896 MW; 72EB67FA2667B398 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 1; Length 574;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ALGVKMA 23  
 |||||  
 Db 430 ALGVKMA 436

RESULT 24

ID MM09\_RABIT STANDARD; PRT; 707 AA.

AC P41246;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE 92 kDa type IV collagenase precursor (EC 3.4.24.35) (92 kDa gelatinase) (Matrix metalloproteinase-9) (MMP-9) (Gelatinase B) (GELB).  
 GN MMP9.  
 OS *Oryctolagus cuniculus* (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NCBI\_TaxId=9986;

RN SEQUENCE FROM N.A.

RP STRAIN-JAPANESE WHITE; TISSUE-Bone;

RX MEDLINE-94253056; PubMed-8195136;

RA Tezuka K.I., Nemoto K., Tezuka Y., Sato T., Ikeda Y., Kobori M., Kawashima H., Eguchi H., Hakeda Y., Kamegawa M.;  
 RT "Identification of matrix metalloproteinase 9 in rabbit osteoclasts.";  
 RL J. Biol. Chem. 269:15006-15009(1994).

RL J. Biol. Chem. 269:15006-15009(1994).

RN SEQUENCE OF 1-171 FROM N.A.

RP STRAIN-NEW ZEALAND WHITE; TISSUE-Liver;

RX MEDLINE-95050662; PubMed-7961810;

RA Fini M.E., Bartlett J.D., Matsubara M., Rinehart W.B., Mody M.K., Girard M.T., Rainville M.;  
 RT "The rabbit gene for 92-kDa matrix metalloproteinase. Role of Ap1 and AP2 in cell type-specific transcription.";  
 RL J. Biol. Chem. 269:28620-28628(1994).

CC -1- FUNCTION: COULD PLAY A ROLE IN BONE OSTEOCLASTIC RESORPTION.

CC -1- CATALYTIC ACTIVITY: Cleavage of gelatin types I and V and collagen types IV and V.

CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.

CC -1- TISSUE SPECIFICITY: OSTEOCLASTS.

CC -1- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN BINDS GLATIN.

CC -1- SIMILARITY: CONTAINS 1 HEMOPLEXIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC METALLOPROTEASE) ALSO KNOWN AS MATRILIXIN SUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: D26514; BAA05520.1; -

DR EMBL: L36050; AAA64358.1; -





```

RESULT 26
ID CUL6_CAEEL STANDARD; PRT; 729 AA.
AC 021346;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE CUL-6 protein.
GN CUL-6 OR KOBE7.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RL Smyle R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z77666; CAB01230.1; -
CC MornPep; KOBE7.7; CEL1928.
CC InterPro; IPR001373; Cullin.
CC Pfam; PF00888; Cullin; 1.
CC SMART; SM00182; CULLIN; 1.
CC PROSITE; PS01256; CULLIN_1; 1.
CC PROSITE; PS50069; CULLIN_2; 1.
DR PROSITE; PS50069; CULLIN_2; 1.
DR PROSITE; PS50069; CULLIN_2; 1.
SQ SEQUENCE 729 AA; 84685 MW; 3D400932DE78E2CF CRC64;

Query Match
Best Local Similarity 6.3%; Score 7; DB 1; Length 729;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AVONTVE 102
Db 651 AVONTVE 657

RESULT 27
ID PRO6_YEAST STANDARD; PRT; 899 AA.
AC P19735;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pre-mRNA splicing factor PRP6.
DE PRP6 OR RNA6 OR YBR055C OR YBR0508.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6288C / GRP88;
RX MEDLINE=90360988; PubMed=2118103;
RA Legrain P., Choulika A.;
RT "The molecular characterization of PRP6 and PRP9 yeast genes reveals
RT a new cysteine/histidine motif common to several splicing factors.";
RL EMO J. 9:2775-2781(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6288C;
RX MEDLINE=95321020; PubMed=7597852;
RA Aljinovic G., Pohl T.M.;

```

```

RT "Sequence and analysis of 24 kb on chromosome II of Saccharomyces
RT cerevisiae.";
RL Yeast 11:475-479(1995).
CC -1- FUNCTION: INVOLVED IN PRE-MRNA SPLICING.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: TO SOME OTHER PRE-MRNA SPLICING FACTORS.
CC -1- SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53465; CAJ37559.1; -
CC EMBL; Z35924; CAJ84998.1; -
CC EMBL; Z46260; CAJ86398.1; -
CC PIR; S12319; S12319.
CC SGD; S0000259; PRP6.
CC InterPro; IPR003107; HAT.
CC InterPro; IPR001440; TPR.
CC Pfam; PF00515; TPR; 3.
CC SMART; SM00386; HAT; 2.
KW mRNA processing; mRNA splicing; TPR repeat; Repeat; Nuclear protein.
FT REPEAT 365 401 TPR 1.
FT REPEAT 462 495 TPR 2.
FT REPEAT 559 593 TPR 3.
FT REPEAT 668 701 TPR 4.
FT REPEAT 805 838 TPR 5.
SQ SEQUENCE 899 AA; 104228 MW; 3E87F8AF63EAC41 CRC64;

Query Match
Best Local Similarity 6.3%; Score 7; DB 1; Length 899;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LVTYQAL 18
Db 758 LVTYQAL 764

RESULT 28
ID ITGA_HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UK01;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGA11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=10486209;
RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Krissansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human Integrin alpha11 subunit (ITGA11).";
RL Genomics 60:179-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal muscle, and Uterus;
RX MEDLINE=99395147; PubMed=10464311;
RA Velling T., Kusche-Gulberg M., Sejersten T., Gulberg D.;
RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
RT Integrin. A collagen-binding, I domain-containing, beta(1)-associated
RT Integrin alpha-chain present in muscle tissues.";

```

RL J. Biol. Chem. 274:25735-25742(1999).  
 RN [3]  
 RP SEQUENCE OF 954-1188 FROM N.A.  
 RC TISSUE-Fibroblast:  
 RA Andreu N., Estivill X., Escarceller M., Sunmy L.  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1 SUBUNIT: INTEGRIN ALPHA11/BETA-1 IS A RECEPTOR FOR COLLAGEN.  
 CC -1 SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11.  
 CC ASSOCIATES WITH BETA-1.  
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1 TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND  
 CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO  
 CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO  
 CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,  
 CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN  
 CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.  
 CC -1 DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING  
 CC FETAL MUSCLE CELLS (IN VITRO).  
 CC -1 DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1 SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1 SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AAF09681; AAF01258.1;  
 CC EMBL: AF137378; AAD51919.2;  
 CC EMBL: AL333064; CAB94392.1;  
 CC HSSP: P17301; IAOX.  
 CC MIM: 604789;  
 CC InterPro: IPR000413; Integrin\_alpha.  
 CC InterPro: IPR002035; VWFA.  
 CC Pfam: PF01839; FG-GAP; 5.  
 CC Pfam: PF00092; VWFA; 1.  
 CC PRINTS: PRO1185; INTEGRIN.  
 CC PRINTS: PRO0453; VWFADOMAIN.  
 CC SMART: SM00327; VWFA; 1.  
 CC SMART: SM00327; Int\_alpha; 5.  
 CC PROSITE: PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
 CC PROSITE: PS50234; VWFA; 1.  
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Signal; Repeat; Calcium; Magnesium; Polymorphism.  
 FT SIGNAL 1 22  
 FT CHAIN 23 1189  
 FT DOMAIN 23 1142  
 FT TRANSMEM 1143 1165  
 FT DOMAIN 1166 1189  
 FT REPEAT 38 94  
 FT REPEAT 102 163  
 FT DOMAIN 167 345  
 FT REPEAT 359 420  
 FT REPEAT 422 475  
 FT REPEAT 477 537  
 FT REPEAT 539 598  
 FT REPEAT 601 653  
 FT DOMAIN 1154 1162  
 FT DOMAIN 1174 1177  
 FT CA\_BIND 488 496  
 FT CA\_BIND 551 559  
 FT CA\_BIND 613 621  
 FT DISULFID 76 83  
 FT DISULFID 121 139  
 FT DISULFID 129 159  
 FT DISULFID 659 668  
 FT DISULFID 674 729  
 FT DISULFID 781 787

FT DISULFID 881 893  
 FT CARBOHYD 82 82  
 FT CARBOHYD 95 95  
 FT CARBOHYD 291 291  
 FT CARBOHYD 331 331  
 FT CARBOHYD 358 358  
 FT CARBOHYD 449 449  
 FT CARBOHYD 462 462  
 FT CARBOHYD 528 528  
 FT CARBOHYD 642 642  
 FT CARBOHYD 694 694  
 FT CARBOHYD 857 857  
 FT CARBOHYD 894 894  
 FT CARBOHYD 973 973  
 FT CARBOHYD 1032 1032  
 FT CARBOHYD 1040 1040  
 FT VARIANT 433 433  
 FT VARIANT 524 524  
 FT VARIANT 972 972  
 FT VARIANT 1003 1003  
 FT VARIANT 1030 1030  
 FT VARIANT 1094 1094  
 FT SEQUENCE 1189 AA; 133609 MW; 60303C08A4AC052 CRC64;  
 SO  
 Query Match 6.3%; Score 7; DB 1; Length 1189;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 ALLVIAL 12  
 DB 1158 ALLVIAL 1164  
 RESULT 29  
 VCAP\_HSV60 STANDARD; PRT; 1345 AA.  
 AC P17887;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Major capsid protein (MCP).  
 GN U57 OR 4L.  
 OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseoloviruses.  
 OC NCBI\_TaxID=10370;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90112641; PubMed=2153237;  
 RX Lattler E., Lawrence G., Liu M.-Y., Barrell B.G., Arrand J.R.  
 RX "Identification, cloning, and expression of the major capsid protein  
 RX gene of human herpesvirus 6."  
 RL J. Virol. 64:714-722(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90080132; PubMed=2152817;  
 RX Lawrence G.L., Chee M., Craxton M.A., Compels U.A., Honess R.W.,  
 RX Barrell B.G.;  
 RX "Human herpesvirus 6 is closely related to human cytomegalovirus."  
 RL J. Virol. 64:287-299(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95266321; PubMed=7747482;  
 RX Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
 RX Martin M.E., Efsthliou S., Craxton M., Maculay H.A.;  
 RX "The DNA sequence of human herpesvirus-6: structure, coding content,

```

RT      for virion structural proteins."
CC      FBS Lett. 200:317-321(1986)..
CC      -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC      HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC      NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC      -1- CATALYTIC ACTIVITY: HYDROLYSTS OF FOUR PEPTIDE BONDS IN THE VIRAL
CC      PRECURSOR POLYPEPTIDE, COMMONLY WITH ASP OR GLU IN THE P6
CC      POSITION, CYS OR THR IN P1 AND SER OR ALA IN P1'.
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA.
CC      -1- MISCELLANEOUS: THE NONSTRUCTURAL PROTEINS NS1 PRESENTS TWO
CC      ALTERNATIVE CLEAVAGE SITES FOR ITS C-TERMINUS, WHICH MAY DEFINE A
CC      SOLUBLE OR A MEMBRANE-BOUND FORM OF NS1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X07755; CAA30581.1; -
DR      EMBL; X03870; CAA27500.1; -
DR      EMBL; X03870; CAA27501.1; ALT SEQ.
DR      EMBL; X03870; CAA27502.1; ALT SEQ.
DR      EMBL; X03870; CAA27503.1; ALT SEQ.
DR      EMBL; X03871; CAA27505.1; -
DR      PIR; A33776; GNMVTB.
DR      PIR; A24055; A24055.
DR      HSSP; P14336; 15VB.
DR      MEROPS; S07.001; -.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR000069; Flavl_M.
DR      InterPro; IPR001157; Flavl_NS1.
DR      InterPro; IPR000752; Flavl_NS2A.
DR      InterPro; IPR000487; Flavl_NS2B.
DR      InterPro; IPR000208; Flavl_NS5.
DR      InterPro; IPR001122; Flavl_capsid.
DR      InterPro; IPR000336; Flavl_glycoprote.
DR      InterPro; IPR001850; Flavl_helicase.
DR      InterPro; IPR002535; Flavl_propep.
DR      InterPro; IPR01003; Flavl_capsid; 1.
DR      Pfam; PF00869; Flavl_glycoprote; 1.
DR      Pfam; PF00949; Flavl_helicase; 1.
DR      Pfam; PF01004; Flavl_M; 1.
DR      Pfam; PF01005; Flavl_NS2A; 1.
DR      Pfam; PF01002; Flavl_NS2B; 1.
DR      Pfam; PF00972; Flavl_NS5; 1.
DR      Pfam; PF01570; Flavl_propep; 1.
DR      PRODOM; PD001496; Flavl_NS1; 1.
DR      PRODOM; PD001556; Flavl_glycoprote; 1.
KW      Polypeptide. Glycoprotein. Transferase: RNA-directed RNA polymerase;
KW      Core protein. Coat protein. Envelope protein. Hydrolase; Helicase;
KW      Atp-binding; Transmembrane;
FT      INT_MET 1 1
FT      CHAIN 1 112
FT      PROPEP 113 205
FT      CHAIN 206 280
FT      CHAIN 281 776
FT      CHAIN 777 1128
FT      CHAIN 1129 1358
FT      CHAIN 1359 1489
FT      CHAIN 1490 2110
FT      CHAIN 2111 2259
FT      CHAIN 2260 2510
FT      CHAIN 2511 3412
FT      NP_BIND 1688 1695
FT      SITE 177 1182
FT      TRANSMEM 101 112

```

FT TRANSMEM 247 259 POTENTIAL.  
 FT TRANSMEM 266 280 POTENTIAL.  
 FT TRANSMEM 738 751 POTENTIAL.  
 FT DISULFID 283 310 BY SIMILARITY.  
 FT DISULFID 340 396 BY SIMILARITY.  
 FT DISULFID 354 385 BY SIMILARITY.  
 FT DISULFID 372 401 BY SIMILARITY.  
 FT DISULFID 466 570 BY SIMILARITY.  
 FT DISULFID 587 618 BY SIMILARITY.  
 FT CARBOHYD 144 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1228 1228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2447 2447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2466 2466 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 381 381 E -> S (IN REF. 3).  
 FT CONFLICT 850 850 E -> D (IN REF. 3).  
 SQ SEQUENCE 3412 AA; 377976 MM; 0F61CE6DCDC9365 CRC64;

Query Match 6.38; Score 7; DB 1; Length 3412;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VLALLV 15  
 DB 1170 VLALLV 1176

RESULT 31  
 POLG\_TBEVH STANDARD: PRT: 3414 AA.  
 ID 001299;  
 AC 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome Polyprotein (containing: capsid protein C (core protein); Matrix  
 DE protein (Envelope protein M); Major envelope protein E; Nonstructural  
 DE proteins NS1, NS2a, NS2b, NS4a and NS4b; Protease/helicase  
 DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)  
 OS Tick-borne encephalitis virus (strain Hypr) (TBEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OX NCBI\_TaxID-70733;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallner G., Mandl C.W., Ecker M., Holzmann H., Stlaasny K.,  
 RA Kunz C., Heinz F.X.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 3357-3414 FROM N.A.  
 RX MEDLINE-91303656; PubMed-1713858;  
 RA Mandl C.W., Kunz C., Heinz F.X.;  
 RT "Presence of poly(A) in a flavivirus: significant differences between  
 RT the 3' noncoding regions of the genomic RNAs of tick-borne  
 RT encephalitis virus strains";  
 RT J. Virol. 65:4070-4077(1991);  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: HYDROLISTS OF FOUR PEPTIDE BONDS IN THE VIRAL  
 CC PRECURSOR POLYPROTEIN, COMMONLY WITH ASP OR GLU IN THE P6  
 CC POSITION, CYS OR THR IN P1 AND SER OR ALA IN P1.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -1- MISCELLANEOUS: THE NONSTRUCTURAL PROTEINS NS1 PRESENTS TWO  
 CC ALTERNATIVE CLEAVAGE SITES FOR ITS C-TERMINUS, WHICH MAY DEFINE A  
 CC SOLUBLE OR A MEMBRANE-BOUND FORM OF NS1.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U39292; AAB53095.1; -  
 CC EMBL: M76660; AAA47904.1; -  
 CC HSSP: P14336; ISVB.  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR000069; Flavi\_M.  
 CC InterPro: IPR001157; Flavi\_NS1.  
 CC InterPro: IPR000752; Flavi\_NS2a.  
 CC InterPro: IPR000487; Flavi\_NS2b.  
 CC InterPro: IPR000404; Flavi\_NS4a.  
 CC InterPro: IPR001328; Flavi\_NS4b.  
 CC InterPro: IPR002088; Flavi\_NS5.  
 CC InterPro: IPR001122; Flavi\_capsid.  
 CC InterPro: IPR000336; Flavi\_glycoprote.  
 CC InterPro: IPR001850; Flavi\_helicase.  
 CC InterPro: IPR002535; Flavi\_propep.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC Pfam: PF01003; Flavi\_capsid; 1.  
 CC Pfam: PF00869; Flavi\_glycoprot; 1.  
 CC Pfam: PF02832; Flavi\_glycop\_C; 1.  
 CC Pfam: PF00949; Flavi\_helicase; 1.  
 CC Pfam: PF01004; Flavi\_M; 1.  
 CC Pfam: PF00948; Flavi\_NS1; 1.  
 CC Pfam: PF01005; Flavi\_NS2a; 1.  
 CC Pfam: PF01002; Flavi\_NS2b; 1.  
 CC Pfam: PF01350; Flavi\_NS4a; 1.  
 CC Pfam: PF01349; Flavi\_NS4b; 1.  
 CC Pfam: PF00972; Flavi\_NS5; 1.  
 CC Pfam: PF01570; Flavi\_propep; 1.  
 CC Pfam: PF01728; FtsJ; 1.  
 CC Pfam: PF00271; Helicase\_C; 1.  
 CC Pfam: PF001496; Flavi\_NS1; 1.  
 CC ProDom: PD001556; Flavi\_glycoprote; 1.  
 CC SMART: SM00490; HelicC; 1.  
 CC PolyProtein: Glycoprotein; Transferrase; RNA-directed RNA polymerase;  
 CC Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;  
 CC ATP-binding; Transmembrane;  
 CC INIT\_MET 1  
 CC FT CHAIN 1 112  
 CC FT PROPEP 113 205  
 CC FT CHAIN 206 280  
 CC FT CHAIN 281 776  
 CC FT CHAIN 777 1128  
 CC FT CHAIN 1129 1358  
 CC FT CHAIN 1359 1489  
 CC FT CHAIN 1490 2110  
 CC FT CHAIN 2111 2259  
 CC FT CHAIN 2260 2511  
 CC FT CHAIN 2512 3414  
 CC FT NP\_BIND 1688 1695  
 CC FT SITE 1779 1782  
 CC FT TRANSMEM 101 112  
 CC FT TRANSMEM 247 259  
 CC FT TRANSMEM 266 280  
 CC FT TRANSMEM 738 751  
 CC FT DISULFID 283 310  
 CC FT DISULFID 340 396  
 CC FT DISULFID 354 385  
 CC FT DISULFID 372 401  
 CC FT DISULFID 466 570  
 CC FT DISULFID 587 618  
 CC FT CARBOHYD 144 434  
 CC CARBOHYD 434 434

ENVELOPE GLYCOPROTEIN M.  
 MAJOR ENVELOPE PROTEIN E.  
 NONSTRUCTURAL PROTEIN NS1 (OR 1190).  
 NONSTRUCTURAL PROTEIN NS2A (OR 1191).  
 NONSTRUCTURAL PROTEIN NS2B.  
 PROTEASE/HELICASE (NS3).  
 NONSTRUCTURAL PROTEIN NS4A.  
 NONSTRUCTURAL PROTEIN NS4B.  
 RNA-DIRECTED RNA POLYMERASE (NS5).  
 ATP BOX  
 DEAD BOX  
 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).  
 POTENTIAL.  
 POTENTIAL.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1649 1649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2447 2447 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2726 2726 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3414 AA: 378539 MW: EC0B1A5325A08C19 CMC64;  
  
Query Match 6.3%; Score 7; DB 1; Length 3414;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 VIALLVLT 15  
DB 1170 VIALLVLT 1176  
  
RESULT 32  
POLG\_TBEVW STANDARD; PRT; 3414 AA.  
AC P14336; Q88493;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [contains: Capsid protein C (Core protein); Matrix  
DE protein (Envelope protein M); Major envelope protein E; Nonstructural  
DE proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase  
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)  
DE (NS5)].  
OS Tick-borne encephalitis virus (Western subtype) (TBEV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Flavivirus.  
OX NCBI\_TaxID=11088;  
RN [1]  
RP SEQUENCE FROM N.A., AND REVISIONS.  
RC STRAIN-NEUDOERFL;  
RA MEDLINE=96036491; PubMed=7483260;  
RA Wallner G., Mandl C.W., Kunz C., Heinz F.X.;  
RT "The flavivirus 3'-noncoding region: extensive size heterogeneity  
RT independent of evolutionary relationships among strains of tick-borne  
RT encephalitis virus.";  
RL Virology 213:169-178(1995).  
RN [2]  
RP SEQUENCE OF 1-779 FROM N.A.  
RC STRAIN-NEUDOERFL;  
RA MEDLINE=8832870; PubMed=3413985;  
RA Mandl C.W., Heinz F.X., Kunz C.;  
RT "Sequence of the structural proteins of tick-borne encephalitis virus  
RT (western subtype) and comparative analysis with other flaviviruses.";  
RL Virology 166:197-205(1988).  
RN [3]  
RP SEQUENCE OF 767-3414 FROM N.A.  
RC STRAIN-NEUDOERFL;  
RA MEDLINE=90051080; PubMed=2554575;  
RA Mandl C.W., Heinz F.X., Stoeckl E., Kunz C.;  
RT "Genome sequence of tick-borne encephalitis virus (Western subtype)  
RT and comparative analysis of nonstructural proteins with other  
RT flaviviruses.";  
RL Virology 173:291-301(1989).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 281-680.  
RX MEDLINE=95272700; PubMed=7753193;  
RA Rey F.A., Heinz F.X., Mandl C.W., Kunz C., Harrison S.C.;  
RT "The envelope glycoprotein from tick-borne encephalitis virus at 2-A  
RT resolution.";  
RL Nature 375:291-298(1995).  
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF FOUR PEPTIDE BONDS IN THE VIRAL  
CC PRECURSOR POLYPEPTIDE, COMMONLY WITH ASP OR GLU IN THE P6  
CC POSITION, CYS OR THR IN P1 AND SER OR ALA IN P1'.  
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
CC -----  
CC EMBL: U27495; AAA6870.1; -  
CC PIR: A31052; GMYVNE.  
CC PDB: 1SVB; 10-JUN-96.  
CC MEROPS: S07.001; -  
CC InterPro: IPR001410; DEAD.  
CC InterPro: IPR000069; Flavi\_M.  
CC InterPro: IPR001157; Flavi\_NS1.  
CC InterPro: IPR000752; Flavi\_NS2A.  
CC InterPro: IPR000487; Flavi\_NS2B.  
CC InterPro: IPR000404; Flavi\_NS4A.  
CC InterPro: IPR001528; Flavi\_NS4B.  
CC InterPro: IPR000208; Flavi\_NS5.  
CC InterPro: IPR001122; Flavi\_Capsid.  
CC InterPro: IPR000336; Flavi\_glycoprote.  
CC InterPro: IPR001850; Flavi\_helicase.  
CC InterPro: IPR002535; Flavi\_propep.  
CC InterPro: IPR002877; Flavi\_NS1.  
CC InterPro: IPR001650; Helicase\_C.  
CC Pfam: PF01003; Flavi\_Capsid; 1.  
CC Pfam: PF00869; Flavi\_glycoprop; 1.  
CC Pfam: PF02832; Flavi\_glycop\_C; 1.  
CC Pfam: PF00949; Flavi\_helicase; 1.  
CC Pfam: PF01004; Flavi\_M; 1.  
CC Pfam: PF00948; Flavi\_NS1; 1.  
CC Pfam: PF01005; Flavi\_NS2A; 1.  
CC Pfam: PF01002; Flavi\_NS2B; 1.  
CC Pfam: PF01350; Flavi\_NS4A; 1.  
CC Pfam: PF01349; Flavi\_NS4B; 1.  
CC Pfam: PF00972; Flavi\_NS5; 1.  
CC Pfam: PF01570; Flavi\_NS5; 1.  
CC Pfam: PF01578; Flavi\_propep; 1.  
CC Pfam: PF01728; Flavi\_C; 1.  
CC Pfam: PF00271; Helicase\_C; 1.  
CC ProDom: PD001496; Flavi\_NS1; 1.  
CC ProDom: PD001556; Flavi\_glycoprote; 1.  
CC SMART: SM00490; Helicase; 1.  
CC PolyProtein: Glycoprotein; Transferase: RNA-directed RNA polymerase;  
CC Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;  
CC ATP-binding; Transmembrane; Nonstructural protein; 3D-structure.  
CC REMOVED FROM CAPSID PROTEIN C BY THE  
CC INIT\_MET 1 1  
FT CHAIN 1 112  
FT PROPEP 113 205  
FT CHAIN 113 205  
FT CHAIN 206 280  
FT CHAIN 281 776  
FT CHAIN 777 1128  
FT CHAIN 1129 1358  
FT CHAIN 1359 1489  
FT CHAIN 1490 2110  
FT CHAIN 2110 2259  
FT CHAIN 2259 2511  
FT CHAIN 2511 3414  
FT CHAIN 3414 391  
FT DOMAIN 391 1779  
FT NP\_BIND 1688 1695  
FT SITE 1779 1782  
FT TRANSMEM 101 112  
FT TRANSMEM 247 259  
FT TRANSMEM 266 280  
  
ENVELOPE GLYCOPROTEIN M.  
MAJOR ENVELOPE PROTEIN E.  
NONSTRUCTURAL PROTEIN NS1.  
NONSTRUCTURAL PROTEIN NS2A.  
NONSTRUCTURAL PROTEIN NS2B.  
NONSTRUCTURAL PROTEIN NS2B.  
PROTEASE/HELICASE (NS3).  
NONSTRUCTURAL PROTEIN NS4A.  
NONSTRUCTURAL PROTEIN NS4B.  
RNA-DIRECTED RNA POLYMERASE (NS5).  
INVOLVED IN FUSION.  
ATP (POTENTIAL).  
DEAD BOX.  
HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).  
POTENTIAL.

```

FT TRANSMEM 738 751 POTENTIAL.
FT DISULFID 283 310
FT DISULFID 340 396
FT DISULFID 354 385
FT DISULFID 372 401
FT DISULFID 466 570
FT DISULFID 587 618
FT CARBOHYD 144 144
FT CARBOHYD 434 434
FT CARBOHYD 861 861
FT CARBOHYD 983 983
FT CARBOHYD 999 999
FT CARBOHYD 2447 2447
SQ SEQUENCE 3414 AA: 378316 MW: 35DBCE014B310B79 CRC64;

Query Match 6.3%; Score 7; DB 1; Length 3414;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 LVALLV 15
DB 1170 LVALLV 1176

RESULT 33
ITF_HUMAN STANDARD; PRT; 80 AA.
AC 007654;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interstitial trefoil factor precursor (HPL.B).
GN TFF3 OR ITF OR TFI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine; and Uterus;
RX MEDLINE=93348192; PubMed=8346203;
RA Hauser F., Poulsom R., Chinery R., Rogers L.A., Hanby A.M.,
RA Wright N.A., Hoffmann W.;
RT "hpl.B, a human p-domain peptide homologous with rat interstitial
RT trefoil factor, is expressed also in the ulcer-associated cell
RT lineage and the uterus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6961-6965(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97224512; PubMed=9070946;
RA Seib T., Bin N., Hilgert K., Seifert M., Theisinger B., Engel M.,
RA Dooley S., Zang K.D., Welter C.;
RT "The three human trefoil genes TFF1, TFF2, and TFF3 are located
RT within a region of 55 kb on chromosome 21q22.3.";
RL Genomics 40:200-202(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Matanabe H., Yada T.,
RA Pak H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,
RA Reithwald K., Rump A., Schilhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Negamine K., Mitsuyma S., Antonarakis S.E.,
RA Mitsuhashi S., Shimizu N., Nordstiek G., Horstischer K., Brandt P.,
RA Schafte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rlesseimann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Leirach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).

```

---

```

RN [4]
RP SEQUENCE OF 8-80 FROM N.A.
RX MEDLINE=93203271; PubMed=8454642;
RA Podolsky D.K., Lynch-Devaney K., Stow J.L., Oates P., Murgue B.,
RA Debeaumont M., Sands B.E., Mahida Y.R.;
RT "Identification of human intestinal trefoil factor. Goblet
RT cell-specific expression of a peptide targeted for apical
RT secretion.";
RL J. Biol. Chem. 268:6694-6702(1993).
CC -1- FUNCTION: MAY HAVE A ROLE IN PROMOTING CELL MIGRATION (MOTOGEN).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED BY GOBLET CELLS OF SMALL AND LARGE
CC -1- INTERSTIAL EPITHELIA AND ALSO BY THE UTERUS.
CC -1- SIMILARITY: CONTAINS 1 P-TYPE (TREFOL) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L15203; AAA59981.1;
DR EMBL: U25657; AAA83628.1;
DR EMBL: U25654; AAA83628.1; JOINED.
DR EMBL: U25656; AAA83628.1; JOINED.
DR EMBL: L08044; AAA36766.1;
DR EMBL: AP001746; BA95531.1; ALT_INT.
DR PIR: A48284; A48284.
DR HSSP: P01359; 2PSP.
DR MIM: 600633;
DR InterPro: IPR000519; P_trefol.
DR Pfam: PF00088; trefol; 1.
DR PRINTS: SM00018; P; 1.
DR PROSITE: PS00025; P_TREFOL; 1.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 80 INTERSTIAL TREFOL FACTOR.
FT DOMAIN 31 72 P-TYPE.
FT DISULFID 32 58 BY SIMILARITY.
FT DISULFID 42 57 BY SIMILARITY.
FT DISULFID 52 69 BY SIMILARITY.
FT CONFLICT 74 76 OEA -> TRKT (IN REF. 4).
SQ SEQUENCE 80 AA: 8641 MW: 9E117A58C0342013 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALL 13
DB 11 LVALL 16

RESULT 34
SECC_AOUAE STANDARD; PRT; 100 AA.
AC 066505;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-export membrane protein secc.
GN SECC OR AO 098.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;

```

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus".  
 RL Nature 392:353-358(1998).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. PARTICIPATES IN A EARLY  
 CC EVENT OF PROTEIN TRANSLLOCATION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SEC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000673; AAC06462.1; -  
 CC Protein transport; Translocation; Transmembrane; Complete proteome.  
 CC TRANSMEM 2 22 POTENTIAL.  
 CC TRANSMEM 53 73  
 CC SEQUENCE 100 AA; 10464 MW; C7183A51B27FA21A CRC64;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LVALL 13  
 DB 63 LVALL 68

RESULT 35  
 VNST\_BUNYW STANDARD: PRT; 101 AA.  
 AC P16494;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Nonstructural protein NS-S.  
 OS Bunyamwera virus.  
 OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Bunyavirus.  
 OX NCBI\_TaxID=11573;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89279304; PubMed=2732714;  
 RA Elliott R.M.;  
 RT "Nucleotide sequence analysis of the small (S) RNA segment of  
 RT Bunyamwera virus, the prototype of the family Bunyaviridae.";  
 RL J. Gen. Virol. 70:1281-1285(1989).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D00353; BAA00262.1; -  
 CC PIR: B31365; MNVUBV.  
 CC PIR: JTO422; JTO422.  
 CC InterPro: IPR000797; Bunya\_NSS.  
 CC Pfam: PF01104; Bunya\_NS-S; 1.  
 CC ProDom: PD002170; Bunya\_NSS; 1.  
 CC Nonstructural protein.  
 CC SEQUENCE 101 AA; 11024 MW; 823C9F75448C6368 CRC64;  
 SQ

Query Match 5.4%; Score 6; DB 1; Length 101;

Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 GLVMTT 85  
 DB 27 GLVMTT 32

RESULT 36  
 VNST\_MAGV STANDARD: PRT; 101 AA.  
 AC P16605;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Nonstructural protein NS-S.  
 OS Maguari virus.  
 OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Bunyavirus.  
 OX NCBI\_TaxID=11575;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89348010; PubMed=2527439;  
 RA Elliott R.M., McGregor A.;  
 RT "Nucleotide sequence and expression of the small (S) RNA segment of  
 RT Maguari bunyavirus.";  
 RL Virology 171:516-524(1989).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M28380; AAA57148.1; -  
 CC EMBL: D13783; BAA02927.1; -  
 CC PIR: B33076; MNVUBM.  
 CC PIR: JTO425; JTO425.  
 CC InterPro: IPR000797; Bunya\_NSS.  
 CC Pfam: PF01104; Bunya\_NS-S; 1.  
 CC ProDom: PD002170; Bunya\_NSS; 1.  
 CC Nonstructural protein.  
 CC SEQUENCE 101 AA; 11099 MW; B1160497B04252DD CRC64;  
 SQ

Query Match 5.4%; Score 6; DB 1; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 GLVMTT 85  
 DB 27 GLVMTT 32

RESULT 37  
 KACA\_RAT STANDARD: PRT; 106 AA.  
 ID KACA\_RAT  
 AC P01836;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain C region, A allele.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DA;  
 RX MEDLINE=82082587; PubMed=6273908;  
 RA Sheppard H.W., Gutman G.A.;  
 RT "Allelic forms of rat kappa chain genes: evidence for strong

FT selection at the level of nucleotide sequence.;

RT Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).

DR PIR: A02116; KAPPA.

DR HSSP: P01842; 2MCS.

DR InterPro: IPR003106; IG\_MHC.

DR InterPro: IPR003597; IG\_C1.

DR Pfam: PF00047; IG\_1.

DR SMART: SM00407; IGc1.1.

DR PROSITE: PS00290; IG\_MHC.1.

KW Immunoglobulin domain; Immunoglobulin C region.

FT NON\_TER 1

FT DISULFID 26 86

FT DISULFID 106 106

FT SEQUENCE 106 AA: 11732 MW; B7E120D9700DD66 CRC64;

Query Match

Best Local Similarity 100.0%; Score 6; DB 1; Length 106;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 LSTRV 51

DB 71 LSTRV 76

RESULT 38

INS\_BRARE STANDARD; PRT; 108 AA.

AC 073727;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Insulin precursor.

GN INS.

OS Brachydanio rerio (zebrafish) (zebra danio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

OC Cypriniformes; Cyprinidae; Danio.

OX NCBI\_Taxid=7955;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98151045; PubMed=9492081;

RA Milewski W.M., Duguay S.J., Chan S.J., Steiner D.F.;

RT "Conservation of PDX-1 structure, function, and expression in zebrafish."

RT zebrafish."

RL Endocrinology 139:1440-1449(1998).

CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-slb.ch/announce/> or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

CC -----

DR EMBL: AF036326; AAC41261.1; -.

DR HSSP: P01308; ILPH.

DR ZFIN: ZDB-GENE-980526-110; Ins.

DR InterPro: IPR00739; Insulin\_IGF\_relaxin.

DR Pfam: PF00049; Insulin.1.

DR PRINTS: PRO0276; INSULINA.

DR PRINTS: PRO0277; INSULINA.

DR SMART: SM00078; IIGF.1.

DR PROSITE: PS00262; INSULIN.1.

KW Insulin family; Hormone; Glucose metabolism; Signal.

FT SIGNAL 1 21

FT CHAIN 22 51

FT PROPEP 54 84

FT CHAIN 88 108

FT DISULFID 30 94

FT DISULFID 42 107

FT DISULFID 93 98

FT SEQUENCE 108 AA: 11904 MW; 31D03CE37BD2C22 CRC64;

Query Match

Best Local Similarity 100.0%; Score 6; DB 1; Length 108;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVL 10

DB 8 GALLVL 13

RESULT 39

DHS\_LACBU STANDARD; PRT; 114 AA.

AC 104193;

DT 20-MAR-1987 (Rel. 04, Created)

DT 20-MAR-1987 (Rel. 04, Last sequence update)

DE Hisidine decarboxylase proenzyme (EC 4.1.1.22) (PI chain) (HDC)

DE (fragment).

OS Lactobacillus buchneri.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

OC Lactobacilli; L1581;

OX NCBI\_Taxid=1581;

RN [1]

RP SEQUENCE.

RX MEDLINE=85131044; PubMed=2857718;

RA Hyatt O.K., Snell E.F.;

RT "Pyruvyl-dependent histidine decarboxylases. Preparation and amino acid sequences of the beta chains of histidine decarboxylase from Clostridium perfringens and Lactobacillus buchneri."

RL J. Biol. Chem. 260:2798-2803(1985).

CC -1- CATALYTIC ACTIVITY: L-histidine -> histamine + CO(2).

CC -1- COFACTOR: PYRUVYL GROUP.

CC -1- SUBUNIT: THE PROENZYME IS A HEXAMER OF IDENTICAL PI CHAINS; EACH PI CHAIN MONOMER IS CLEAVED TO FORM A SMALL (OR BETA) CHAIN AND A LARGE (OR ALPHA) CHAIN BY NONHYDROLYTIC SELF-CATALYSIS.

CC PIR: A01080; DCLABH.

DR HSSP: P00862; LPEA.

DR InterPro: IPR003427; HDC.

DR Pfam: PF02329; HDC.1.

KW Lyase; Decarboxylase; Pyruvate; Zymogen.

FT CHAIN 1 81

FT CHAIN 82 >114

FT SITE 81 82

FT MOD.RES 82 82

FT NON\_TER 114 114

FT SEQUENCE 114 AA: 12490 MW; 95907A08AC761166 CRC64;

Query Match

Best Local Similarity 100.0%; Score 6; DB 1; Length 114;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 KINTIG 110

DB 6 KINTIG 11

RESULT 40

MERT\_SHEPU STANDARD; PRT; 115 AA.

ID MERT\_SHEPU

AC 054462;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)



DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Mercury transport protein (Mercury ion transport protein).  
 GN MERT.  
 OS Shewanella putrefaciens (Pseudomonas putrefaciens).  
 OC Plasmid Inc PMERPH.  
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
 OC Shewanella.  
 OX NCBI\_Taxid=24;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-9730939; PubMed-9167257;  
 RA Osborn A.M., Bruce K.D., Strike P., Ritchie D.A.;  
 RT "Distribution, diversity and evolution of the bacterial mercury  
 resistance (mer) operon";  
 RL FEMS Microbiol. Rev. 19:239-262(1997).  
 CC -1- FUNCTION: INVOLVED IN MERCURIC TRANSPORT. PASSES A HG(2+) ION  
 FROM THE PERIPLASMIC MERT PROTEIN TO THE MERCURIC REDUCTASE  
 (MERA).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z49196; CAA89054.1; -  
 DR InterPro: IPR003457; MERT.  
 DR Pfam: PF02411; MERT; 1.  
 KW Transport; Mercuric resistance; Inner membrane; Mercury; Plasmid;  
 KM Transmembrane.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 45 65 POTENTIAL.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT METAL 23 23 HG(2+) (BY SIMILARITY).  
 FT METAL 24 24 HG(2+) (BY SIMILARITY).  
 FT METAL 75 75 HG(2+) (BY SIMILARITY).  
 FT METAL 81 81 HG(2+) (BY SIMILARITY).  
 SQ SEQUENCE 115 AA; 12555 MW; C4ADC83CD389074E CRC64;

Query Match 5.4%; Score 6; DB 1; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLIAL 12  
 ID LVLIAL 12  
 DB 55 LVLIAL 60

RESULT 41  
 CHH4\_PENMO STANDARD; PRT; 120 AA.  
 AC 097386;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Crustacean hyperglycemic hormone 4 precursor (CHH) (Pm-SGP-IV).  
 GN CHH4.  
 OS Penaeus monodon (Penaeid shrimp).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;  
 OC Penaeidae; Penaeus.  
 OX NCBI\_Taxid=6687;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Pyestalk;  
 RL PubMed-10804243;  
 RA Davey M.L., Hall M.R., Willis R.H., Oliver R.W.A., Thurn M.J.,  
 RA Wilson K.J.;

RT "Five crustacean hyperglycemic family hormones of penaeus monodon:  
 RT complementary DNA sequence and identification in single sinus glands  
 RT by electrospray ionization-Fourier transform mass spectrometry";  
 RL Mar. Biotechnol. 2:80-91(2000).  
 CC -1- FUNCTION: ABUNDANT HORMONE IN THE SINUS GLAND OF ISPODS AND  
 CC DECAPODS WHICH CONTROLS THE BLOOD SUGAR LEVEL. HAS A SECRETAGOGUE  
 CC ACTION OVER THE AMYLASE RELEASED FROM THE MIDGUT GLAND. MAY ACT AS  
 CC A STRESS HORMONE AND MAY BE INVOLVED IN THE CONTROL OF MOLTING AND  
 CC REPRODUCTION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD CHH/MH/GIH/VIH FAMILY OF  
 CC HORMONES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF104389; AAC84145.1; -  
 DR InterPro: IPR001166; CHH\_MH\_GIH.  
 DR Pfam: PF01147; Crust\_neurohorm; 1.  
 DR PRINTS: PR00550; HYPERGLYCEMIC.  
 DR PROSITE: PS01250; CHH\_MH\_GIH; 1.  
 KW Neuropeptide; Hormone; Glucose metabolism; Amidation; Signal;  
 KM Cleavage on pair of basic residues; Multigene family.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT PEPTIDE 27 44 CHH PRECURSOR RELATED PEPTIDE (CPRP).  
 FT PEPTIDE 47 118 CRUSTACEAN HYPERGLYCEMIC HORMONE 4.  
 FT DISULFID 53 89 BY SIMILARITY.  
 FT DISULFID 69 85 BY SIMILARITY.  
 FT DISULFID 72 98 BY SIMILARITY.  
 FT MOD\_RES 118 118 AMIDATION (G-119 PROVIDE AMIDE GROUP) (BY  
 FT SIMILARITY).  
 SQ SEQUENCE 120 AA; 12987 MW; D04F39BDF7C4C1B9 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 83;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIA 11  
 ID ALLVIA 11  
 DB 13 ALLVIA 18

RESULT 42  
 YBAV\_ECOLI STANDARD; PRT; 123 AA.  
 ID YBAV\_ECOLI  
 AC P77415;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ybav precursor.  
 GN YBAV OR B0442.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,

RA Duncan M., Federpiet N., Hyman R., Kalman S., Komp C., Kurd O.,  
 RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.,  
 RA Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / W3110.  
 RA Hatada E., Ohmori H., Qiao Y., Tsuji M., Fukuda R.,  
 RA Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.  
 RL [1]  
 CC -1- SIMILARITY: STRONG, TO H. INFLUENZA H1108.  
 CC -1- SIMILARITY: TO B. SUBTILIS COMEA.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000150; AAC73545.1;  
 CC EMBL: 082664; BAB40198.1;  
 CC EMBL: DB2943; BAB1646.1;  
 CC Ecogene: EG13250; ypaV.  
 CC Interpro: IPR000445; ypaV.  
 CC Interpro: IPR003583; HH1.  
 CC SMART: SM00278; HH1.  
 CC Hypothetical protein; Signal; Complete proteome.  
 FT SIGNAL 1 25  
 FT CHAIN 26 123  
 FT SEQUENCE 123 AA; 12704 MW; 3CE382B0B15A5BC CRC64;  
 SQ  
 Query Match 5.4%; Score 6; DB 1; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 85;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 100 TVEDLK 105  
 DB 99 TVEDLK 104  
 RESULT 43  
 ID RNP\_SPEAH STANDARD; PRT; 125 AA.  
 AC P16414;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Ribonuclease pancreatic (EC 3.1.27.5) (RNase I) (RNase A).  
 GN RNASEI OR RNSI.  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;  
 OC Spalax.  
 OC NCBI\_TaxID=30637;  
 RA [1]  
 RP SEQUENCE.  
 RC TISSUE=Pancreas; PubMed=2673297;  
 RA MEDLINE=89374807; Muebels G., Baintema J.J., Nevo E.;  
 RA Schueler C., Neutboom B., Wuebelts G., Baintema J.J., Nevo E.;  
 RA The amino-acid sequence of pancreatic ribonuclease from the mole rat  
 RA Spalax ehrenbergi, chromosomal species n = 60.  
 RL Biol. Chem. Hoppe-Seyler 370:583-589(1989).  
 CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 3'-  
 CC phosphonucleotides and 3'-phosphonucleotides ending in C-  
 CC for U-P with 2',3'-cyclic phosphate intermediates.  
 CC -1- SUBCELLULAR LOCATION: PANCREAS.  
 CC -1- TISSUE SPECIFICITY: PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.  
 DR PIR: S04503; S04503.  
 DR HSP: P00636; I55B.  
 DR Interpro: IPR001427; RNaseA.  
 DR Pfam: PF00074; rnaaseA.1.

DR PRINTS: PR00794; RIBONUCLEASE.  
 DR PRODOM: PD000535; RNaseA.1.  
 DR SMART: SM00092; RNase-PC; 1.  
 DR PROSITE: PS00127; RNASE\_PANCREATIC; 1.  
 KW Hydrolyase; Endonuclease; Glycoprotein.  
 FT CARBOHYD 35 35  
 FT DISULFID 27 85  
 FT DISULFID 41 96  
 FT DISULFID 59 111  
 FT DISULFID 66 73  
 FT ACT\_SITE 12 12  
 FT ACT\_SITE 42 42  
 FT ACT\_SITE 120 120  
 FT ACT\_SITE 120 120  
 FT SEQUENCE 125 AA; 14031 MW; 9D074E74261E7E4 CRC64;  
 SQ  
 Query Match 5.4%; Score 6; DB 1; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 85 TISSSK 90  
 DB 87 TISSSK 92  
 RESULT 44  
 ID CHMO\_BACSU STANDARD; PRT; 127 AA.  
 AC P19080;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chorismate mutase (EC 5.4.99.5) (Cm).  
 GN AKOH.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 CC NCBI\_TaxID=1423;  
 RA [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-35.  
 RC STRAIN=168 / MARBURG;  
 RA MEDLINE=90148962; PubMed=2105742;  
 RA Gray J.V., Gellinell-Pimpaneau B., Knowles J.R.;  
 RA Monofunctional chorismate mutase from Bacillus subtilis:  
 RA purification of the protein, molecular cloning of the gene, and  
 RA overexpression of the gene product in Escherichia coli.  
 RL Biochemistry 29:376-383(1990).  
 GN [2]  
 OS SEQUENCE FROM N.A.  
 RA Hemmer D.J.;  
 RA Sequence of Bacillus subtilis dbpA, mtr(A,B), gerc(1-3), ndk, cher,  
 RA aro(B,E,F,H), trp(A-F), hsh, and tyrA genes.  
 RA Submitted (JAN-1992) to the EMBL/Genbank/DBJ databases.  
 RL [3]  
 RP STRUCTURE BY NMR.  
 RA MEDLINE=93229495; PubMed=8471608;  
 RA Rajagopalan J.S., Taylor K.M., Jaffe E.K.;  
 RA 13C NMR studies of the enzyme-product complex of Bacillus subtilis  
 RA chorismate mutase.  
 RL Biochemistry 32:3965-3972(1993).  
 RA [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA MEDLINE=93391402; PubMed=8378335;  
 RA Chook Y.M., Ke H., Lipscomb W.N.;  
 RA Crystal structures of the monofunctional chorismate mutase from  
 RA Bacillus subtilis and its complex with a transition state analog.  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8600-8603(1993).  
 RA [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RA MEDLINE=94322390; PubMed=8046752;  
 RA Chook Y.M., Gray J.V., Ke H., Lipscomb W.N.;  
 RA The monofunctional chorismate mutase from Bacillus subtilis:  
 RA structure determination of chorismate mutase and its complexes with a

RT transition state analog and prephenate, and implications for the  
 RT mechanism of the enzymatic reaction."  
 RL J. Mol. Biol. 240:476-500(1994).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).  
 RX MEDLINE-20280181; PubMed-10818343;  
 RA Lader J.E., Reddy P., Davis A., Tordova M., Howard A.J.,  
 RA Gilliland G.L.;  
 RT "The 1.30 A resolution structure of the Bacillus subtilis chorismate  
 RT mutase catalytic homotrimer."  
 RL Acta Crystallogr. D 56:673-683(2000).  
 CC -1- CATALYTIC ACTIVITY: Chorismate -> prephenate.  
 CC -1- ENZYME REGULATION: THIS ENZYME IS MONOFUNCTIONAL, AND ITS ACTIVITY  
 CC IS UNAFFECTED BY THE END-PRODUCT AROMATIC AMINO ACIDS.  
 CC -1- PATHWAY: BRANCH POINT OF THE BIOSYNTHETIC PATHWAY LEADING TO THE  
 CC THREE AROMATIC AMINO ACIDS, PHENYLALANINE, TYROSINE, & TRYPTOPHAN.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: M32278; AAA22249.1; -  
 CC DR EMBL: M80245; AAA20861.1; -  
 CC DR EMBL: 299115; CAB14185.1; -  
 CC DR PIR: A33894; A33894.  
 CC DR PIR: 1COM: 22-JUN-94.  
 CC DR PDB: 1COM: 22-JUN-94.  
 CC DR PDB: 2CHS: 31-JUL-94.  
 CC DR PDB: 2CHT: 31-JUL-94.  
 CC DR PDB: 1DBF: 07-JUN-00.  
 CC DR Subtilist; Bg10286; arOH.  
 CC DR InterPro: IPR002701; Chorismate\_mut.  
 CC DR Pfam: PF01817; Chorismate\_mut; 1.  
 CC DR Aromatic amino acid biosynthesis; Isomerase; 3D-structure;  
 CC KW Complete proteome.  
 CC FT CONFLICT 112 112 A -> V (TN REF. 2).  
 CC SQ SEQUENCE 127 AA; 14489 MW; 8DD03614FE1F403C CRC64;

Query Match 5.4%; Score 6; DB 1; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 DSLTRK 50  
 DB 118 DSLTRK 123

RESULT 45  
 Y070\_TREPA  
 ID Y070\_TREPA STANDARD; PRT; 128 AA.  
 AC 083109;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein TP0070.  
 GN TP0070.  
 OS Treponema pallidum.  
 CC Bacteria: Spirochaetales; Spirochaetaceae; Treponema.  
 CC NCBI\_TaxID=160;  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLS;  
 RX MEDLINE-98332770; PubMed-9665876;  
 RA Frazer C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."  
 RL Science 281:375-388(1998).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: AE001191; AAC65067.1; -  
 CC DR TIGR: TP0070; -  
 CC DR Hypothetical protein: Transmembrane; Complete proteome.  
 CC FT TRANSMEM 2 22 POTENTIAL.  
 CC FT TRANSMEM 34 54 POTENTIAL.  
 CC FT TRANSMEM 64 84 POTENTIAL.  
 CC FT TRANSMEM 108 128 POTENTIAL.  
 CC SQ SEQUENCE 128 AA; 13552 MW; 81B55415B523D5DF CRC64;

Query Match 5.4%; Score 6; DB 1; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVLALL 13  
 DB 42 LVLALL 47

RESULT 46  
 VEB\_COPY  
 ID VEB\_COPY STANDARD; PRT; 144 AA.  
 AC 089808;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE B6 protein.  
 GN B6.  
 CC OS Canine oral papillomavirus (COPY).  
 CC OC Viruses: dsDNA viruses, no RNA stage; Papillomaviridae;  
 CC OC Papillomavirus.  
 CC OX NCBI\_TaxID=35258;  
 CC RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=v62;  
 RA Isegawa N., Ohta M., Shirasawa H., Tokita H., Simizu B., Yamaura A.;  
 RT "Nucleotide sequence of a canine oral papillomavirus containing a long  
 RT noncoding region."  
 RL Int. J. Oncol. 7:155-159(1995).  
 CC -1- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE  
 CC STRANDED DNA (IN VITRO).  
 CC -1- SUBCELLULAR LOCATION: Nuclear matrix-associated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: D55633; BAA09498.1; -  
 CC DR EMBL: L22695; AAA61744.1; -  
 CC DR InterPro: IPR001334; B6.  
 CC DR Pfam: PF00518; B6; 1.  
 CC KW Early protein; DNA-binding; Nuclear protein; Zinc-finger.  
 CC ZN\_FING 33 69 BY SIMILARITY.

FT ZN\_FING 105 140 BY SIMILARITY.  
SQ SEQUENCE 144 AA; 16131 MW; 229D9190A3049B8C CRC64;

Query Match 5.48; Score 6; DB 1; Length 144;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 LLDLST 48  
DB 26 LLDLST 31

## RESULT 47

IL2\_BOVIN STANDARD; PRT; 155 AA.

AC P05016;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).  
GN IL2 OR IL-2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86205869; PubMed=3517854;  
RA Carrell D.P., McKereghan K., Larsen A., Cantrell M.A., Anderson D.,  
RA Gillis S., Cosman D., Baker P. P.;  
RT "Cloning, sequence, and expression of bovine interleukin 2."  
RL Proc. Natl. Acad. Sci. U.S.A. 83:3223-3227(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86205870; PubMed=3488415;  
RA Reeves R., Spies A.G., Nissen M.S., Buck C.D., Weinberg A.D.,  
RA Barr P.J., Magnuson N.S., Magnuson J.A.;  
RT "Molecular cloning of a functional bovine interleukin 2 cDNA."  
RL Proc. Natl. Acad. Sci. U.S.A. 83:3328-3332(1986).  
RN [3]  
RP SEQUENCE OF 1-22 FROM N.A.

RA TISSUE=Thymus;  
RA Aikoeva N.N., Vinogradova T.V., Votoshin O.N.;  
RL Submitted (DEC-1989) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR  
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL  
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE  
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-  
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M12791; AAA30586.1; -;  
DR EMBL: M13204; AAA21143.1; ALT\_INIT.  
DR EMBL: X17201; CAA35062.1; -;  
DR EMBL: X52687; CAA36912.1; -;  
DR HSSP: P01585; 31MK.  
DR InterPro: IPR000779; Interleukin-2.  
DR Pfam: PF00715; IL2; 1.  
DR PRINTS: PR00265; INTERLEUKIN2.  
DR PRODOM: PD003649; Interleukin-2; 1.  
DR SMART: SM00189; IL2; 1.  
DR PROSITE: PS00424; INTERLEUKIN\_2; 1.

KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;  
KW T-cell.  
FT SIGNAL. 1 20 INTERLEUKIN-2.  
FT CHAIN 21 155 BY SIMILARITY.  
FT DISULFID 79 127 O-LINKED (GALNAC... ) (BY SIMILARITY).  
FT CARBOHYD 23 23 V->A (IN REF. 2).  
FT CONFLICT 66 66  
SQ SEQUENCE 155 AA; 17627 MW; 816667DEA052EDF CRC64;

Query Match 5.48; Score 6; DB 1; Length 155;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 KVNATE 55  
DB 68 KVNATE 73

## RESULT 48

IL2\_SHEEP STANDARD; PRT; 155 AA.

AC P19114;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).  
GN IL2  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91016933; PubMed=2216781;  
RA Goodall J.C., Emery D.C., Perry A.C.F., English L.S., Hall L.;  
RT "CDNA cloning of ovine interleukin 2 by PCR."  
RL Nucleic Acids Res. 18:5883-5883(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91088336; PubMed=2263496;  
RA Seow H.F., Rothel J.S., Radford A.J., Wood P.R.;  
RT "The molecular cloning of ovine interleukin 2 gene by the polymerase  
RT chain reaction."  
RL Nucleic Acids Res. 18:7175-7175(1990).  
RN [3]  
RP SEQUENCE OF 21-153 FROM N.A.  
RX MEDLINE=95367627; PubMed=7543777;  
RA Bujdoso R., Williamson M.L., Roy D., Hunt P., Blacklows B., Sargan D.,  
RA McConnell I.;  
RT "Molecular cloning and expression of DNA encoding ovine interleukin  
RT 2."  
RL Cytokine 7:223-231(1995).

CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR  
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL  
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE  
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-  
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X53934; CAA37881.1; -;  
DR EMBL: X55641; CAA39165.1; -;  
DR EMBL: X60148; CAA42722.1; -;

DR EMBL: A19169; CAA01448.1; -  
 DR PIR: S11488; S11488.  
 DR PIR: S13102; S13102.  
 DR PIR: S15517; S15517.  
 DR HSP: P01585; 31NK.  
 DR InterPro: IPR00779; Interleukin-2.  
 DR Pfam: PF00715; IL2; 1.  
 DR PRINTS: PR00265; INTERLEUKIN2.  
 DR ProDom: PD003649; Interleukin-2; 1.  
 DR SMART: SM00189; IL2; 1.  
 DR PROSITE: PS00424; INTERLEUKIN\_2; 1.  
 DR Cytokine; Glycoprotein; Immune response; Signal; Growth factor;  
 T-cell.  
 KW SIGNAL.  
 FT CHAIN 1 20 INTERLEUKIN-2.  
 FT CARBOHYD 23 23 O-LINKED (GALNAC. .) (BY SIMILARITY).  
 FT DISULFID 79 127 BY SIMILARITY.  
 FT CONFLICT 6 6 P -> L (IN REF. 2).  
 SQ SEQUENCE 155 AA; 17661 MW; 3EAA6D67667494D CRC64;

Query Match 5.4%; Score 6; DB 1; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 KVNATE 55  
 |||||  
 DB 68 KVNATE 73

RESULT 49  
 ID RR7\_GINBI STANDARD; PRT; 155 AA.  
 AC 09GFL3:  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S7.  
 GN RPS7.  
 OS Ginkgo biloba (Ginkgo).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Ginkgophyta; Ginkgoales; Ginkgoaceae; Ginkgo.  
 OX NCBI\_TaxID=311;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11080123;  
 RA Graham S.W., Olmstead R.G.;  
 RT "Utility of 17 chloroplast genes for inferring the phylogeny of the  
 RT basal angiosperms.";  
 RL Am. J. Bot. 87:1712-1730(2000).  
 CC -1- FUNCTION: Protein S7 binds specifically to part of the 3' end of  
 CC 16S ribosomal RNA (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF123779; AAC26116.1; -  
 DR InterPro: IPR000235; Ribosomal\_S7.  
 DR Pfam: PF00177; Ribosomal\_S7; 1.  
 DR ProDom: PD000817; Ribosomal\_S7; 1.  
 DR PROSITE: PS00052; RIBOSOMAL\_S7; 1.  
 DR PROSITE: PS00052; RIBOSOMAL\_S7; 1.  
 KW Ribosomal protein; rRNA-binding; Chloroplast.  
 SQ SEQUENCE 155 AA; 17782 MW; 18155A6FC446F369 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 155;

Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 60 AMKKIQ 65  
 |||||  
 DB 46 AMKKIQ 51

RESULT 50  
 ID RR7\_LOTJA STANDARD; PRT; 155 AA.  
 AC 09B1A1:  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S7.  
 GN RPS7-A AND RPS7-B.  
 OS Lotus japonicus.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.  
 OX NCBI\_TaxID=34305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Accession MG-20;  
 RX MEDLINE=21082929; PubMed=11214967;  
 RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;  
 RT "Complete structure of the chloroplast genome of a legume, Lotus  
 RT japonicus.";  
 RL DNA Res. 7:323-330(2000).  
 CC -1- FUNCTION: Protein S7 binds specifically to part of the 3' end of  
 CC 16S ribosomal RNA (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AF002983; BAB33254.1; -  
 DR EMBL: AP002983; BAB33240.1; -  
 DR InterPro: IPR000235; Ribosomal\_S7.  
 DR Pfam: PF00177; Ribosomal\_S7; 1.  
 DR ProDom: PD000817; Ribosomal\_S7; 1.  
 DR PROSITE: PS00052; RIBOSOMAL\_S7; 1.  
 KW Ribosomal protein; rRNA-binding; Chloroplast.  
 SQ SEQUENCE 155 AA; 17374 MW; 7D7AC7D7DF860CBF CRC64;

Query Match 5.4%; Score 6; DB 1; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 AMKKIQ 65  
 |||||  
 DB 46 AMKKIQ 51

RESULT 51  
 ID RR7\_SOYBN STANDARD; PRT; 155 AA.  
 AC P07135;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S7.  
 GN RPS7.  
 OS Glycine max (Soybean).  
 OG Chloroplast.

CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 CC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;  
 CC eurosids I: Fabales; Fabaceae: Papilionoideae; Phaseoleae; Glycine.  
 CC NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=67174761; PubMed=3562230;  
 RA von Allmen J.M., Stutz E.;  
 RT Complete sequence of 'divided' rps12 (r-protein S12) and rps7 (r-  
 RT protein S7) gene in soybean chloroplast DNA.";  
 RT Nucleic Acids Res. 15:2387-2387(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. MAPLE ARROW;  
 RA von Allmen J.M., Stutz E.;  
 RT Submitted (MAY-1988) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: X05013; CAA28662.1;  
 DR EMBL: X07675; CAA30523.1;  
 DR PIR: S05718; R3S7.  
 DR PIR: B26744; B2674.  
 DR HSSP: P1291; TRS.  
 DR Mendel; 4823; Gluma: rps7.1.  
 DR InterPro: IPR000235; Ribosomal\_S7.  
 DR Pfam: PF00177; Ribosomal\_S7.1.  
 DR ProDom: PD000817; Ribosomal\_S7.1.  
 DR PROSITE: PS00052; RIBOSOMAL\_S7.1.  
 KW Ribosomal protein; rRNA-binding; Chloroplast.  
 SQ SEQUENCE 155 AA; 17361 MW; 73943FB50B942E2P CRC64;  
 -----  
 Query Match 5.4%; Score 6; DB 1; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 60 AMKKIQ 65  
 DB 46 AMKKIQ 51  
 -----  
 RESULT 52  
 ID VES\_RHPV1 STANDARD: PRT; 157 AA.  
 AC P24834;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Probable E5 protein.  
 GN E5.  
 OS Rhesus papillomavirus type 1 (Rhpv 1).  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10570;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91135018; PubMed=1847267;  
 RA Ostrow R.S., Labresh K.V., Farias A.J.;  
 RT "Characterization of the complete Rhpv 1 genomic sequence and an  
 RT integration locus from a metastatic tumor.";  
 RT Virology 181:424-429(1991).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: M60184; AAA79315.1; ALT\_SEQ.  
 DR EMBL: M60184; AAA79316.1; ALT\_SEQ.  
 DR PIR: F38503; W5WLR1.  
 KW Early protein.  
 SQ SEQUENCE 157 AA; 17398 MW; AC7AA67158844686 CRC64;  
 -----  
 Query Match 5.4%; Score 6; DB 1; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 9 VVALVY 14  
 DB 140 VVALVY 145  
 -----  
 RESULT 53  
 ID YRN5\_CAEEL STANDARD: PRT; 160 AA.  
 AC O09419;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 16.4 kDa protein R07B1.5 in chromosome X precursor.  
 GN R07B1.5.  
 OS Caenorhabditis elegans.  
 CC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditidae;  
 CC Rhabditidae; peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Kershaw J.;  
 RT Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: SOME, TO C.ELEGANS C30G12.4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL: Z48621; CAA88542.1;  
 DR WormPep: R07B1.5; CE01631.  
 DR InterPro: IPR002601; C6.  
 DR Pfam: PF01681; C6.1.  
 KW Hypothetical protein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 160  
 FT DOMAIN 22 58  
 SQ SEQUENCE 160 AA; 16413 MW; E9C21E2219712A13 CRC64;  
 -----  
 Query Match 5.4%; Score 6; DB 1; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 33 FFAVAN 38  
 DB 12 FFAVAN 17  
 -----  
 RESULT 54  
 ID PHAB\_AGLINE STANDARD: PRT; 161 AA.  
 AC P28556;  
 -----

```

CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION. THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: U14682; AAA61733.1; -.
CC DR HSSP: P01585; 3INK.
CC DR InterPro: IPR000779; Interleukin-2.
CC DR Pfam: PF00715; IL2; 1.
CC DR PRINTS: PR00265; INTERLEUKIN2.
CC DR ProDom: PD003649; Interleukin-2; 1.
CC DR SMART: SM00189; IL2; 1.
CC DR PROSITE: PS00424; INTERLEUKIN_2; 1.
CC KW CYTOKINE; Glycoprotein; Immune response; Signal; Growth factor;
CC T-cell.
CC FT SIGNAL 1 20 BY SIMILARITY.
CC FT CHAIN 21 162 INTERLEUKIN-2.
CC FT DISULFID 29 134 BY SIMILARITY.
CC FT CARBOHYD 73 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
CC FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 162 AA; 18389 MW; 6B572799B1B6E799 CRC64;
CC -----
CC Query Match 5.4%; Score 6; DB 1; Length 162;
CC Best Local Similarity 100.0%; Pred. No. 1;le+02;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 50 KVNATE 55
Db 68 KVNATE 73
-----
RESULT 56
LVCV_BPP21
ID LVCV_BPP21 STANDARD; PRT; 165 AA.
AC P27359;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysozyme (EC 3.2.1.17) (Lysis protein) (Mucamidase) (Endolysin).
GN R.
OS Bacteriophage P21 (Bacteriophage 21).
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda phage group.
OX NCBI_Taxid=10711;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91210180; PubMed=2019562;
RA Bonovich M.T., Young R.;
RT "Dual start motif in two lambdaoid S genes unrelated to lambda S.";
RL J. Bacteriol. 173:2897-2905(1991).
CC -1- FUNCTION: ESSENTIAL FOR LYSIS OF BACTERIAL CELL WALL. BY SHOWING
CC CELL WALL HYDROLYZING ACTIVITY. ACTS AS A TRANSGLYCOSYLASE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4'-beta-linkages between N-
CC acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC heteropolymers of the prokaryotes cell walls.
CC -1- SIMILARITY: BELONGS TO FAMILY 24 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M65239; AAA32350.1; -  
 DR PIR: S22906; IZBP21  
 DR InterPro: IPR002196; Phage\_lysozyme.  
 DR Pfam: PF00959; Phage\_lysozyme.1  
 KM Late protein; Hydrolyse; Glycosidase; Bacteriolytic enzyme.  
 FT ACT\_SITE 35 35 PROTON DONOR (BY SIMILARITY).  
 FT NUCLEOPHILE 44 44 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 165 AA; 17996 MW; 14ECCB83232D3C CRC64;

Query Match 5.4%; Score 6; DB 1; Length 165;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TMRGAL 7  
 DB 97 TMRGAL 102

RESULT 57  
 ID SULA\_SERMA STANDARD; PRT; 168 AA.  
 AC P08845;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE Cell division inhibitor.  
 GN SULA  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OK Serratia.  
 RN NCBL\_TaxID-615;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87248093; PubMed-3297925;  
 RA Friedl R., Braun G., Honore N., Cole S.T.;  
 RT "Evolution of the enterobacterial sula gene: a component of the SOS  
 RT system encoding an inhibitor of cell division."  
 RT Gene 52:31-40(1987).  
 CC -1- FUNCTION: SULA IS A COMPONENT OF THE SOS SYSTEM AND AN INHIBITOR  
 CC OF CELL DIVISION. ACCUMULATION OF SULA CAUSES RAPID CESSATION OF  
 CC CELL DIVISION AND THE APPEARANCE OF LONG, NON-SEPTATE FILAMENTS.  
 CC THE EXPRESSION OF SULA IS RERESSED BY LEXA PROTEIN. FTSZ SEEMS  
 CC TO BE THE TARGET OF SULA.  
 CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE.  
 CC -1- MISCELLANEOUS: THE N-TERMINUS OF SULA PROTEIN MIGHT BE INVOLVED IN  
 CC RECOGNIZING THE CELL DIVISION APPARATUS.  
 CC -1- SIMILARITY: THE C-TERMINUS SHOWS SIMILARITY TO THE N PROTEIN OF  
 CC PHAGE LAMBDA  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M16468; AAA26582.1; -  
 DR PIR: D29016; D29016.  
 DR Cell division; Septation; SOS response; Inner membrane.  
 KM DOMAIN 31 149 CONSERVED REGION  
 FT SIMILAR 150 168 TO N PROTEIN OF PHAGE LAMBDA  
 FT DOMAIN 160 168 LOW PROTEIN RECOGNITION SITE (PROBABLE).  
 SQ SEQUENCE 168 AA; 19116 MW; EB681DE3CF59FD11 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 168;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 70 ENGLIS 75  
 DB 30 ENGLIS 35

RESULT 58  
 ID PHAF\_CYAPA STANDARD; PRT; 169 AA.  
 AC P48087;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Allophycocyanin B18 subunit.  
 GN APCF.  
 OS Cyanophora paradoxa.  
 OC Cyanelle.  
 CC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.  
 CX NCBL\_TaxID-2762;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LB555 / PRINGSHEIM.  
 RA Bryant D.A., Michalowski C.B., Loeffelhardt W., Bohmert H.J.,  
 RA "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa."  
 RA Plant Mol. Biol. Rep. 13:327-332(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LB555 / PRINGSHEIM;  
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,  
 RA Farley J.Y., Schluchter W.M., Chung S., Nemann-Spallart C.,  
 RA Steiner J.M., Jakowitsch J., Bohmert H.J., Bryant D.A.;  
 RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:  
 RT the genetic complexity of a primitive plastid."  
 RT (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,  
 RL Schwemmler W. (eds.);  
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg  
 RL (1997).  
 CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BLUE PIGMENT-PROTEIN  
 CC FROM THE PHCOBILIPROTEIN COMPLEX. ALLOPHYCOCYANIN HAS A MAXIMUM  
 CC ABSORPTION AT APPROXIMATELY 650 NANOMETERS.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.  
 CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U30821; AAA81307.1; -  
 DR HSSP: P00318; 1B33.  
 DR Mendel: 7667; CYAPA:apcf.1.  
 DR InterPro: IPR001659; Phycobillsome.  
 DR Pfam: PF00502; Phycobillsome; 1.  
 KM Phycobillsome; Electron transport; Photosynthesis; Blue pigment;  
 KW Cyanelle; Methylation.  
 FT MOD.RES 72 72 METHYLATION (BY SIMILARITY).  
 FT BINDING 82 82 PHYCOCYANOBILIN CHROMOPHORE  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 169 AA; 18380 MW; 37D47A57479790AA CRC64;

Query Match 5.4%; Score 6; DB 1; Length 169;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 76 RVLDGL 81  
 DB 111111





```

DE Holiday Junction DNA helicase RUVA.
GN RUVA OR TM0165.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Matk J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999)
CC -1- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP REPAIRS
CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC HOMOLOGOUS RECOMBINATION. RUVA IS AN HELICASE THAT MEDIATES THE
CC HOLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC REANNEALING. RUVA STIMULATES, IN THE PRESENCE OF DNA, THE WEAK
CC ATPASE ACTIVITY OF RUVB (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A COMPLEX WITH RUVB.
CC -1- SIMILARITY: BELONGS TO THE RUVA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001701; AAD35258.1;
DR HSSP: P40832; 18VS.
DR TIGR: TM0165;
DR InterPro: IPR003583; HHH_1.
DR InterPro: IPR000085; RUVA.
DR Pfam: PF01330; RUVA_1.
DR Pfam: PF02904; RUVA_1.
DR Pfam: PF008268; RUVA_1.
DR PRODOM: PD008268; RUVA_1.
DR SMART: SM00278; HHH1; 2.
DR DNA repair; SOS response; DNA-binding; DNA recombination; Helicase;
KW Complete proteome.
KW SEQUENCE 188 AA; 20818 MW; 28E7F39D67FD61DD CRC64;
SQ

```

```

Query Match          5.4%; Score 6; DB 1; Length 188;
Best local similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 46 LSLTKV 51
   |||||
DB 73 LSLTKV 78

```

```

RESULT 62
RNH2_SYNY3
ID RNH2_SYNY3 STANDARD; PRT; 190 AA.
AC P72657;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease H1I (Ec 3.1.26.4) (RNase H1I).
GN RNHB OR SR1130.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugliura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K., Okumura S.,
RA Shilpm S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: Degrades THE RIBONUCLEOTIDE MOIETY ON RNA-DNA HYBRID
CC MOLECULES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC PHOSPHOMONONESTER.
CC -1- Cofactor: MANGANESE (BY SIMILARITY).
CC -1- PHOSPHOMONONESTER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RNASE H1I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90899; BAA16659.1;
DR InterPro: IPR001352; RNase_H1I.
DR Pfam: PF01351; RNase_H1I.
KW Hydrolyase; Nuclease; Endonuclease; Manganese; Complete proteome.
FT ACT SITE 5 5 BY SIMILARITY.
FT ACT SITE 101 101 BY SIMILARITY.
FT ACT SITE 120 120 BY SIMILARITY.
FT ACT SITE 120 120 BY SIMILARITY.
SQ SEQUENCE 190 AA; 20684 MW; CB745958228BBA6D CRC64;

```

```

Query Match          5.4%; Score 6; DB 1; Length 190;
Best local similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 16 QALGVK 21
   |||||
DB 32 QALGVK 37

```

```

RESULT 63
YK25_YEAST
ID YK25_YEAST STANDARD; PRT; 191 AA.
AC P36138;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 21.1 kDa protein in GAP1-NAP1 intergenic region.
GN YKR045C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Wisner S., Ustretarazu L.A., Jauniaux J.-C.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBS databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z28270; CAA82121.1;
DR PIR: S38117; S38117.

```

DR SGD: S0001753; YKR045C.  
 KW Hypothetical protein.  
 SW SEQUENCE 191 AA; 21110 MM; 409C95BBD418D1B9 CRC64;  
 Query Match 5.4%; Score 6; DB 1; Length 191;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 84 TTTSSS 89  
 Db 67 TTTSSS 72  
 RESULT 64  
 BID\_HUMAN STANDARD; PRT; 195 AA.  
 AC P55957;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE BH3 interacting domain death agonist (BID).  
 GN BID.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eultheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97078762; PubMed=8918887;  
 RA Wang K., Yin X.-M., Chao D.T., Millman C.L., Korsmeyer S.J.;  
 RT "BID: a novel BH3 domain-only death agonist.";  
 RL Genes Dev. 10:2859-2869(1996).  
 RT duplicated in cat eye syndrome chromosomes and to mouse chromosome 6.";  
 RL Genomics 51:472-475(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M.,  
 RA Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M.,  
 RA Parsons J., Rifkin L., Kohling T., Soares M., Tan F.,  
 RA Trevisakis E., Waterston R., Williamson A., Wohlmann P., Wilson R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-110 FROM N.A.  
 RA Fujiwara T., Hirano H., Hishigaki H., Horie M., Kawai A., Kuga Y.,  
 RA Kyushiki H., Nagata M., Okuno S., Ozaki K., Shimizu F.,  
 RA Kyumada Y., Shinomiya H., Suzuki M., Takaichi A., Takeda S.,  
 RA Watanabe T., Maekawa H., Nakamura Y., Takahashi E.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 1-74.  
 RX MEDLINE=96159527; PubMed=8593609;  
 RA Trostler J.A., Long K.R., Murrell J.R., Stotler C.J.,  
 RA Gusella J.F., Buckler A.J.;  
 RT "An expression-independent catalog of genes from human chromosome 22.";  
 RL Genome Res. 5:214-224(1995).  
 RN [6]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=99189747; PubMed=10089877;  
 RA Chou J.J., Li H., Salvesen G.S., Yuan J., Wagner G.;  
 RT "Solution structure of BID, an intracellular amplifier of apoptotic signaling.";  
 RL Cell 96:615-624(1999).  
 CC -1- FUNCTION: INDUCES ICE-LIKE PROTEASES AND APOPTOSIS. COUNTERS THE PROTECTIVE EFFECT OF BCL-2 (BY SIMILARITY).

CC -1- SUBUNIT: FORMS HETERODIMERS EITHER WITH THE PRO-APOPTOTIC PROTEIN BAX OR THE ANTI-APOPTOTIC PROTEIN BCL-2 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY CYTOPLASMIC (BY SIMILARITY).  
 CC -1- DOMAIN: INTACT BH3 DOMAIN IS REQUIRED BY BIK, BID, BAX, BAD AND BAX FOR THEIR PRO-APOPTOTIC ACTIVITY AND FOR THEIR INTERACTION WITH ANTI-APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOGY DOMAIN 3 (BH3).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL: AF042083; AAC34365.1; -;  
 CC EMBL: H23042; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: R09650; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: R09537; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: C17308; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: H55483; -; NOT\_ANNOTATED\_CDS.  
 CC PDB: 2BID; 27-JAN-00.  
 CC MIM: 601997; -;  
 CC InterPro: IPR000712; BCL\_2.  
 CC PROSITE: PS01259; BH3; 1.  
 CC Apoptosis; 3D-structure.  
 CC FT DOMAIN 86 100 BH3.  
 CC SQ SEQUENCE 195 AA; 21994 MM; B17A07334C1AFBEF CRC64;  
 QY 8 LVLALL 13  
 Db 149 LVLALL 154  
 Query Match 5.4%; Score 6; DB 1; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 65  
 LEPL\_SYNY3  
 ID LEPL\_SYNY3 STANDARD; PRT; 196 AA.  
 AC P72660;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable signal peptidase I-1 (PC 3.4.21.89) (Spase I-1) (Leader peptidase I-1).  
 DE LEPL OR SLI0716.  
 GN LEPL OR SLI0716.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,  
 RA Shinochi T., Matsuno A., Muraiki A., Nakazaki N., Natsu K., Okumura S.,  
 RA Hosono S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF N-TERMINAL LEADER SEQUENCES FROM SECRETED PROTEINS. PRECURSOR.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S26; ALSO KNOWN AS TYPE I LEADER PEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: D90899; BAAL662.1; -  
CC HSSP; P00803; IB12.  
CC InterPro: IPR000508; Peptidase\_S26.  
CC InterPro: IPR000223; Peptidase\_S26A.  
CC Pfam: PF00461; Peptidase\_S26; 2.  
CC PRINTS: PR00727; LEADERPTASE.  
CC PROSITE: PS00501; SPASE\_1.1; 1.  
CC PROSITE: PS00760; SPASE\_1.2; 1.  
CC PROSITE: PS00761; SPASE\_1.3; 1.  
CC Transmembrane: Hydrolyase; Protease; Complete proteome.  
CC DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 17 35 POTENTIAL.  
CC DOMAIN 36 196 PERIPLASMIC (POTENTIAL).  
CC ACT\_SITE 44 44 BY SIMILARITY.  
CC ACT\_SITE 94 94 BY SIMILARITY.  
CC SEQUENCE 196 AA; 22229 MW; EB02533BF7341573 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALL 13  
DB 24 LVALL 29

RESULT 66  
YKGB\_ECOLI STANDARD; PRT; 197 AA.  
AC P75685; P71303;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ykgb.  
GN YKGB OR B0301  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia

CC NCBI\_TaxID=562;  
CC SEQUENCE FROM N.A.  
CC STRAIN-K12 / MG1655;  
CC MEDLINE-9742617; PubMed-9278503;  
CC Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
CC Mau B., Shao Y.;  
CC "The complete genome sequence of Escherichia coli K-12.";  
CC Science 277:1453-1474(1997).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
CC Federapfel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
CC Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
CC Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1 SIMILARITY: TO H. INFLUENZAE HI0219.  
CC -1 SIMILARITY: TO P. HAEMOLYTICA HYPOTHETICAL PROTEIN IN PURT/MVAI  
CC 5/REGION (AC P46928)  
CC -1 CAUTION: REP. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 138.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC use by non-profit institutions as long as its content is in no way

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: AE000137; AAC73404.1; ALT-INIT.  
CC EMBL: D73857; AAB18029.1; ALT\_FRAME.  
CC EcoGene; EG13579; ykgb.  
CC Hypothetical protein; Transmembrane; Complete proteome.  
CC TRANSMEM 16 36 POTENTIAL.  
CC TRANSMEM 86 106 POTENTIAL.  
CC TRANSMEM 113 133 POTENTIAL.  
CC SEQUENCE 197 AA; 21899 MW; FC45E94197045C5E CRC64;

Query Match 5.4%; Score 6; DB 1; Length 197;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVLA 11  
DB 101 ALLVLA 106

RESULT 67  
AIS\_ECOLI STANDARD; PRT; 200 AA.  
AC P45565; P77314;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE AIS protein.  
GN AIS OR B2252.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia

CC NCBI\_TaxID=562;  
CC SEQUENCE FROM N.A.  
CC STRAIN-K12;  
CC MEDLINE-9742617; PubMed-9278503;  
CC Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
CC Mau B., Shao Y.;  
CC "The complete genome sequence of Escherichia coli K-12.";  
CC Science 277:1453-1474(1997).  
CC [3]  
CC SEQUENCE FROM N.A.  
CC STRAIN-K12;  
CC MEDLINE-97349980; PubMed-9205837;  
CC Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
CC Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsubashi N.,  
CC Oshibuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
CC Oshima T., Oyama S., Saito N., Sampel G., Satoh Y., Sivasubraman S.,  
CC Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
CC Yamagata S., Horinouchi T.;  
CC "Construction of a contiguous 874-kb sequence of the Escherichia coli  
CC K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
CC RT analysis of its sequence features.";  
CC DNA Res. 4:91-113(1997).  
CC -1 INDUCTION: BY ALBUMINUM.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X83874; CAA58754.1; -  
 DR EMBL: AE000315; AAC75312.1; -  
 DR EMBL: D90856; CAB21999.1; -  
 DR Ecocyc: EGI3155; a1s.  
 KM Transmembrane: Complete proteome.  
 FT TRANSMEM 14 33 POTENTIAL.  
 FT CONFLICT 200 200 H->Y (IN REF. 1).  
 SQ SEQUENCE 200 AA; 22256 MW; A210382654AC6B35 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 78 LDGLVM 83  
 DB 179 LDGLVM 184

RESULT 68  
 PSS\_RHLP  
 ID PSS\_RHLP STANDARD; PRT; 200 AA.  
 AC P10498;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-MAY-1992 (Rel. 22, Last annotation update)  
 DE Exopolysaccharide production protein PSS.  
 GN PSS.  
 OS Rhizobium leguminosarum (biovar phaseoli).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=385;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8002;  
 RX MEDLINE=89127136; PubMed=2851702;  
 RA Borthakur D., Barker R.F., Latchford J.W., Rossen L., Johnston A.W.B.;  
 RT "Analysis of pss genes of Rhizobium leguminosarum required for  
 RT exopolysaccharide synthesis and nodulation of peas: their primary  
 RT structure and their interaction with psi and other nodulation  
 RT genes.";  
 RL Mol. Gen. Genet. 213:155-162(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X12568; CAA31079.1; -  
 DR PIR: S03811; S03811.  
 DR InterPro: IPR003362; Bact\_transf.  
 DR Pfam: PF02397; Bact\_transf. 1.  
 KM Exopolysaccharide synthesis; Nodulation.  
 SQ SEQUENCE 200 AA; 22318 MW; 26A7ECE9193A0A8 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVLA 11  
 DB 17 ALLVLA 22

RESULT 69

GCIP\_RANPI  
 ID GCIP\_RANPI STANDARD; PRT; 205 AA.  
 AC 073763;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Guanylyl cyclase inhibitory protein.  
 GN GCIP.  
 OS Rana pipiens (Northern leopard frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 OX NCBI\_TaxID=8404;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Retina;  
 RX MEDLINE=98206755; PubMed=9546678;  
 RA Li N., Fariss R.N., Zhang K., Otto-Bruc A.E., Haeseleer F.,  
 RA Bronson J.D., Qin N., Yamazaki A., Subbaraya I., Millam A.H.,  
 RA Palczewski K., Baehr W.;  
 RT "Guanylate-cyclase-inhibitory protein is a frog retinal Ca2+-binding  
 RT protein related to mammalian guanylate-cyclase-activating proteins.";  
 RL Eur. J. Biochem. 252:591-599(1998).  
 CC -I- FUNCTION: DOES NOT STIMULATE GUANYLYL CYCLASE (GC) WHEN FREE  
 CC CALCIUM ION CONCENTRATION IS LOW, BUT INHIBITS GC WHEN FREE  
 CC -I- TISSUE SPECIFICITY: RETINA; INNER SEGMENTS, SOMATA AND SYNAPTIC  
 CC TERMINALS OF CONE RECEPTORS.  
 CC -I- MISCELLANEOUS: BINDS TWO CALCIUM IONS (BY SIMILARITY).  
 CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF047884; AF047884.1; -  
 DR HSSP: P51177; 1JBA.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR001125; Recoverin.  
 DR Pfam: PF00036; ehand. 3.  
 DR PRINTS: PR00450; RECOVERIN.  
 DR SMART: SM00054; Eph. 3.  
 DR PROSITE: PS00018; EF\_HAND; 2.  
 KM Calcium-binding; Repeat; Vision; Myristate.  
 KW K  
 FT INIT\_MET 0  
 FT LIPID 1  
 FT DOMAIN 27 38 MYRISTATE (POTENTIAL).  
 FT CA\_BIND 63 74 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).  
 FT CA\_BIND 99 110 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 147 158 EF-HAND 3 (POTENTIAL).  
 FT ANCESTRAL CALCIUM SITE 4 (POTENTIAL).  
 SQ SEQUENCE 205 AA; 23521 MW; 61830977AFED40F1 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SLTKVN 52  
 DB 126 SLTKVN 131

RESULT 70  
 AMIS\_RHOER  
 ID AMIS\_RHOER STANDARD; PRT; 206 AA.  
 AC 053185;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Putative transporter protein am152.

GN	AMIS2	Rhodococcus erythropolis.
OS	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.	
CC	NCR1_TextID=18933; [1]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN-BREVIARCTARIUM SP. / STRAIN R312.	
RC	MEDLINE-97136714. Pubmed-8982091.	
RX	Chebreou H., Bigey F., Arnard A., Galzy P.;	
RT	"amide metabolism: a putative ABC transporter in Rhodococcus sp.	
RL	R312.";	
CC	Gene 182:215-218(1996).	
CC	-1- FUNCTION: POSSIBLE TRANSPORTER THAT MIGHT BE RESPONSIBLE FOR THE	
CC	ADSORPTION OF AMIDASE SUBSTRATES OR RELEASE OF THEIR HYDROLYSIS	
CC	PRODUCTS.	
CC	-1- SIMILARITY: BELONGS TO THE AMIS/DREI FAMILY.	
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).	
CC	CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	CC	
DR	EMBL; 246523; CAA86569.1; -	
DR	InterPro; IPR003211; AMIS_Drei.	
DR	Pfam; PF02293; AMIS_Drei; 1	
KM	Transmembrane; Transport.	
FT	TRANSMEM 4 24 POTENTIAL.	
FT	TRANSMEM 29 49 POTENTIAL.	
FT	TRANSMEM 56 76 POTENTIAL.	
FT	TRANSMEM 86 106 POTENTIAL.	
FT	TRANSMEM 113 133 POTENTIAL.	
FT	TRANSMEM 142 162 POTENTIAL.	
FT	TRANSMEM 193 POTENTIAL.	
SQ	SEQUENCE 206 AA; 21645 MW; A/3BE2D54F9B4663 CRC64;	
Oy	5 GALLVYL 10 	
Db	179 GALLVL 184	
RESULT 71		
ID	TRPE_KLUJA STANDARD:	
AC	P13997; PRT: 210 AA.	
DT	01-JAN-1990 (Rel. 13, Created)	
DT	01-JAN-1990 (Rel. 13, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).	
GN	TRPI.	
OS	Kluyveromyces fragilis (Yeast).	
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.	
CC	NCR1_TextID=28985; [1]	
RN	SEQUENCE FROM N.A.	
RP	MEDLINE-89189093; PubMed-2538971;	
RX	Stark M.J.R., Milner J.S.;	
RT	"Cloning and analysis of the Kluyveromyces fragilis TRPI gene: a	
RT	chromosomal locus flanked by genes encoding inorganic pyrophosphatase	
RL	and histone H3.";	
CC	Yeast 5:35-50(1989).	
CC	-1- CATALYTIC ACTIVITY: N-(5'-phospho-beta-D-ribose-1)-anthranilate - 1-	
CC	(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate.	

```

CC -1- PATWAY: THIRD STEP IN TRYPTOPHAN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE TRP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X14230; CAA32445.1; -
DR PIR: S07893; ISYKNI.
DR HSSP: O56320; IDL3.
DR InterPro: IPR001240; PRA1.
DR Pfam: PF00697; PRA1; 1.
DR Isomerase; Tryptophan biosynthesis.
SQ SEQUENCE 210 AA; 22880 MW; 0F406A05E83CFA95 CRC64;
OY 41 ELLDLD 46
Db 116 ELLDLD 121
-----
RESULT 72
YN68_DEIRA STANDARD; PRT: 215 AA.
AC Q9RWM7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein DR2368.
GN DR2368.
OC Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OC [1]_TaxID=1239;
OC [1]_
PC SEQUENCE FROM N.A.
RA MEDLINE-20036896; PubMed-10567266;
FX White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Holt D., Gilm M.L., Nelson K.C., Richardson D.L.,
RA Moffitt K.S., Qin H., Jiang L., Penhale M., Cosby K., Shen H.,
RA Mavetyan J., Lam P., McDonald L., Utterback T., Wasth H.,
RA Mekarova K.S., Aravind L., Daly M.J., Minor K.W., Fleischman R.D.,
RA Fraser C.M., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA "genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE UPF0126 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002067; NAP11915.1; -
DR TIGR: DR2368; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 15 POTENTIAL.
FT TRANSMEM 39 POTENTIAL.
FT TRANSMEM 59 POTENTIAL.
FT TRANSMEM 75 POTENTIAL.
FT TRANSMEM 101 POTENTIAL.
FT TRANSMEM 121 POTENTIAL.

```

FT TRANSMEM 123 143 POTENTIAL.  
 FT TRANSMEM 162 182 POTENTIAL.  
 SO SEQUENCE 215 AA: 22928 MW: B127F94CF6109279 CRC64;

Query Match  
 Best local Similarity 100.0%; Score 6; DB 1; Length 215;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVL 10  
 DB 190 GALLVL 195

## RESULT 73

HYPE\_ECOLI  
 ID HYPE\_ECOLI STANDARD: PRT; 216 AA.  
 AC P77524;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hydrogenase-4 component E (EC 1.-.-.-).  
 GN HYPE OR B2485 OR Z3745 OR ECS3347.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562, 83334;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA Andrews S.C., McClay J., Ambler A., Quail M., Berks B.C., Guest J.R.,  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97349980; PubMed-9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horinouchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE-21074935; PubMed-11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN (5)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / RIMD 050952;  
 RX MEDLINE-21156231; PubMed-11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: M63654; AAB88567.1; -;  
 CC DR EMBL: AE000335; AAC75538.1; -;  
 CC DR EMBL: D90877; BAA16373.1; -;  
 CC DR EMBL: AE005478; AAG57595.1; -;  
 CC DR EMBL: AP002561; BAB36770.1; -;  
 CC DR Ecogene; EG14213; hyfe.  
 CC KM Oxidoreductase; Transmembrane; Inner membrane; Complete proteome.  
 CC FT TRANSMEM 4 24 POTENTIAL.  
 CC FT TRANSMEM 39 59 POTENTIAL.  
 CC FT TRANSMEM 60 80 POTENTIAL.  
 CC FT TRANSMEM 93 113 POTENTIAL.  
 CC FT TRANSMEM 123 143 POTENTIAL.  
 CC FT TRANSMEM 151 171 POTENTIAL.  
 CC FT TRANSMEM 176 196 POTENTIAL.  
 SO SEQUENCE 216 AA: 23361 MW: 141A76B9961DBF4F CRC64;

Query Match  
 Best local Similarity 100.0%; Score 6; DB 1; Length 216;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALL 13  
 DB 164 LVALL 169

## RESULT 74

ID Y700\_RICPR STANDARD: PRT; 221 AA.  
 AC Q9ZCM4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical ABC transporter ATP-binding protein RP700.  
 GN RP700.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE-99039499; PubMed-9823893;  
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,  
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria.";  
 RL Nature 396:133-140(1998).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. LIGD SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ235272; CAI5136.1; -  
 DR InterPro: IPR003593; AAA.  
 DR InterPro: IPR003439; ABC\_transport.  
 DR InterPro: IPR001687; ATP\_A.  
 DR Pfam: PF00005; ABC\_tran. 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 DR Hypothetical protein: ATP-binding; Transport; Complete proteome.  
 FT NP\_BIND 42 49 ATP (BY SIMILARITY).  
 SQ SEQUENCE 221 AA; 24846 MW; B465A4E7896E03D5 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 RLVLDT 81  
 |||||  
 Db 23 RLVLDT 28

RESULT 75  
 ID NK13\_RAT STANDARD; PRT; 223 AA.  
 AC P27471.  
 DT 01-AUG-1992 (Rel. 23, Created).  
 DT 01-AUG-1992 (Rel. 23, Last sequence update).  
 DT 01-JUN-1994 (Rel. 29, Last annotation update).  
 DE Natural killer cell surface protein P1-3.2.3 (NKR-P1 3.2.3) (Antigen  
 DE 3.2.3).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:90378305; PubMed:2339464;  
 RA Giordano R., Rudert W.A., Vavassori C., Chambers W.H.,  
 RA Hissrodt J.C., Trucco M.;  
 RT "NKR-P1" a signal transduction molecule on natural killer cells.\*;  
 RT Science 249:1298-1300(1990).  
 CC - FUNCTION: MEDIATES TRANSMEMBRANE SIGNALING IN NATURAL KILLER  
 CC (NK) CELLS AND SO MAY ACT AS A RECEPTOR ABLE TO SELECTIVELY  
 CC TRIGGER NK CELL ACTIVITY.  
 CC - SUBUNIT: HOMODIMER.  
 CC - SUBCELLULAR LOCATION: Type II membrane protein.  
 CC - TISSUE SPECIFICITY: NATURAL KILLER CELLS.  
 CC - MISCELLANEOUS: LIGAND BINDING MAY BE CALCIUM DEPENDENT.  
 CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M62891; AAA1710.1; -  
 DR PIR: A35917; A35917.  
 DR InterPro: IPR001304; Lectin\_C.  
 DR Pfam: PF00059; Lectin\_C; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR PROSITE: PS00615; C\_TYPE\_LECTIN\_1; FALSE\_NEG.  
 DR PROSITE: PS50041; C\_TYPE\_LECTIN\_2; 1.  
 KW Glycoprotein; Antigen; Transmembrane; Signal-anchor; Lectin.  
 FT DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 44 63 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT

FT DOMAIN 64 223 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 93 212 C-TYPE LECTIN (LONG FORM).  
 FT DISULFID 94 105 BY SIMILARITY.  
 FT DISULFID 122 210 BY SIMILARITY.  
 FT DISULFID 189 202 BY SIMILARITY.  
 FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 223 AA; 24551 MW; FCD12B212DDF4330 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLVAL 12  
 |||||  
 Db 51 LVLVAL 56

RESULT 76  
 ID YB13\_MYCPN STANDARD; PRT; 223 AA.  
 AC P75449.  
 DT 16-OCT-2001 (Rel. 40, Created).  
 DT 16-OCT-2001 (Rel. 40, Last sequence update).  
 DT 16-OCT-2001 (Rel. 40, Last annotation update).  
 DE Hypothetical protein MPN113 (C09\_orf23).  
 GN MPN113 OR MP041.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE:97105885; PubMed:8948633;  
 RA Himmelfrich R., Hilbert H., Plogens H., Pirkl E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RT pneumoniae".  
 RT Nucleic Acids Res. 24:4420-4449(1996).  
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000005; AAB95689.1; -  
 DR Hypothetical protein; Transmembrane; Complete proteome.  
 KW TRANSMEM 28 48 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 88 108 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 176 196 POTENTIAL.  
 SQ SEQUENCE 223 AA; 24406 MW; 4066156893CE018C CRC64;

Query Match 5.4%; Score 6; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 LALLVT 15  
 |||||  
 Db 187 LALLVT 192

RESULT 77  
 VG2R\_AMEPV



ID VG2R\_AMEPV STANDARD; PRT; 225 AA.  
AC P29818;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Hypothetical G2R protein.  
GN G2.  
OS Amsacta moorei entomopoxvirus (AmePV).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
OC Entomopoxvirus B.  
OC NCBI\_TaxID=28321;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92046310; PubMed=1942245;  
RA Hall R.L., Moyer R.W.;  
RT "Identification, cloning, and sequencing of a fragment of Amsacta  
RT moorei entomopoxvirus DNA containing the spheroïdin gene and three  
RT vaccinia virus-related open reading frames.";  
RL J. Virol. 65:6516-6527(1991).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
CC EMBL: M77182; AAA42380.1; -  
CC DR PIR: B41561; W2VZG2.  
CC KM Hypothetical protein.  
CC SO SEQUENCE 225 AA; 26008 MW; FBBEAA3DD5D4F37B CRC64;

Query Match 5.4%; Score 6; DB 1; Length 225;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLYTQA 17  
Db 75 LLYTQA 80

RESULT 78  
RUYA\_TREPA STANDARD; PRT; 227 AA.  
ID RUYA\_TREPA  
AC O83554;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Holliday junction DNA helicase ruva.  
GN RUYA OR TP0543.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OC NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=NICHOLS;  
RX MEDLINE=98332770; PubMed=9665876;  
RX Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardson J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Kralak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,  
RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RT Science 281:375-388(1998).  
-----  
CC -!- FUNCTION: THE RUYA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATRES  
CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,  
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN  
CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE

CC HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND  
CC REANNEALING. RUYA STIMULATES, IN THE PRESENCE OF DNA, THE WEAK  
CC ATPASE ACTIVITY OF RUVB (BY SIMILARITY).  
CC -!- SUBUNIT: FORMS A COMPLEX WITH RUVB (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE RUYA FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
CC EMBL: AE001229; AAC65527.1; -  
CC DR HSSP: P08576; ICUK.  
CC DR TIGR: TP0543; -  
CC DR InterPro: IPR003583; HHH\_1.  
CC DR InterPro: IPR000085; RUYA.  
CC DR Pfam: PF01330; RUYA.1.  
CC DR Pfam: PF02904; RUYA\_II.1.  
CC DR PRODOM: PD006268; RUYA; 1.  
CC DR SMART: SM00278; HHH1; 2.  
CC KM DNA repair: SOS response; DNA-binding; DNA recombination; Helicase;  
CC Complete proteome.  
CC SO SEQUENCE 227 AA; 24399 MW; 6B66933517D37156 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 227;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LSLTKV 51  
Db 73 LSLTKV 78

RESULT 79  
TOX3\_BORPE STANDARD; PRT; 227 AA.  
ID TOX3\_BORPE  
AC P04979;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Pertussis toxin subunit 3 precursor (PTX S3) (Islet-activating  
DE protein S3) (IAP S3).  
GN PTXC.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OC NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BP165;  
RX MEDLINE=86259651; PubMed=2873570;  
RX Nicotia A., Perugini M., Franzini C., Casaghi M.C., Borri M.G.,  
RA Antoni G., Almoni M., Neri P., Ratti G., Rappuoli R.;  
RT "Cloning and sequencing of the pertussis toxin genes: operon  
RT structure and gene duplication.";  
RT Proc. Natl. Acad. Sci. U.S.A. 83:4631-4635(1986).  
-----  
CC -!- FUNCTION: THE RUYA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATRES  
CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,  
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN  
CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE

RL Structure 2:45-57(1994).  
RM X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE-96219180; PubMed-8637000.  
RN Hedges B., Boonhoo A., Cockle S.A., Read R.J.,  
RT "Crystal" structure of the pertussis toxin-ATP complex: a molecular  
J. Mol. Biol. 258:661-671(1996).  
CC -1- FUNCTION: PTA OLIGOMER B BINDS TO RECEPTORS ON THE EUKARYOTIC CELL  
CC SUPRACE AND INDUCES THE TRANSLLOCATION OF THE TOXIC SUBUNIT  
CC ACROSS THE CELL MEMBRANE.  
CC -1- SUBUNIT: PERTUSSIS TOXIN CONTAINS FIVE DIFFERENT CHAINS, S1-S5,  
CC THEY ARE ORGANIZED INTO 2 FUNCTIONAL SUBUNITS: A, COMPOSED OF S1  
CC S4 (B BINDS TO THE MEMBRANE RECEPTORS). DIMERS OF S2-S4 AND S3-S4  
CC ARE HELD TOGETHER BY S5.  
CC -1- SIMILARITY: STRONG, TO SUBUNIT S2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M14378; AAA83984.1; -  
CC EMBL: M13223; AAA23895.1; -  
CC EMBL: A13359; CA01095.1; -  
CC PIR: C24394; MEBR31  
CC PDB: 1PPT; 26-JAN-95  
CC PDB: 1BCD; 05-JUN-95  
CC PDB: 1PTO; 15-SEP-95  
CC InterPro: IPR003699; Harport toxB.  
CC Pfam: PF02918; Pertussis toxin; 1.  
CC DR PRINTS: PR01366; BORPPTOXIN; 1.  
CC DR Membrane; Toxin; Signal; Whooping cough; 3D-structure.  
CC KW CHAIN 1 227  
CC FT SIGNAL 26  
CC FT CHAIN 26 227  
CC FT DISULFID 51 115  
CC FT DISULFID 148 162  
CC FT DISULFID 220 227  
CC SO SEQUENCE 227 AA; 24988 MW; 8E15948065BBA87D CRC64;  
  
Query Match 5.4%; Score 6; DB 1; Length 227;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 8 LVYAL 13  
DB 16 LVYAL 21  
  
RESULT 80  
COX2\_LUMTE STANDARD; PRT; 228 AA.  
ID COX2\_LUMTE  
AC Q37545;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
GN COXI.  
OS Lumbricus terrestris (Common earthworm).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
OC Lumbricina; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
OC NCBL\_Taxid-6398;  
RX NCBI\_Taxid-6398;  
RN SEQUENCE FROM N.A.  
RP MEDLINE-96042914; PubMed-8536978;  
RA Boore J.L., Brown W.M.;  
RT "Complete sequence of the mitochondrial DNA of the annelid worm

RL Lumbricus terrestris;  
RM Genetics 141:305-319(1995).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINDING COPPER  
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.  
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) -> 4 ferrocyclochrome  
CC c + 2 H(2)O.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
CC inner membrane.  
CC  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U24570; AAC46851.1; -  
CC HSP: P08306; IAR1  
CC InterPro: IPR001505; COX2.  
CC InterPro: IPR002429; Cyt\_c-ox\_2.  
CC DR Pfam: PF00116; COX2; 1.  
CC DR Pfam: PF02790; COX2\_TM; 1.  
CC DR PRINTS: PR01166; CYCOXIDASEII.  
CC DR PRODOM: PD000131; COX2; 1.  
CC DR PROSITE: PS00078; COX2; 1.  
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
CC Electron transport; Respiratory chain.  
CC FT DOMAIN 1 26  
CC FT TRANSFM 47 26  
CC FT DOMAIN 48 60  
CC FT TRANSFM 61 81  
CC FT DOMAIN 82 228  
CC FT DOMAIN 161 161  
CC FT METAL 196 196  
CC FT METAL 200 200  
CC FT METAL 204 204  
CC SO SEQUENCE 228 AA; 25659 MW; A1CE3834F3BD80D3 CRC64;  
  
Query Match 5.4%; Score 6; DB 1; Length 228;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 LVYAL 12  
DB 73 LVYAL 78  
  
RESULT 81  
COX2\_SCYCA STANDARD; PRT; 230 AA.  
ID COX2\_SCYCA  
AC Q79404;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
GN COXI OR COII.  
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Galeomorphi; Galeidae; Carcharhiniformes;  
OC Scyliorhinidae; Scyliorhinus.  
RX NCBI\_Taxid-7830;  
RN NCBI\_Taxid-7830;  
RP SEQUENCE FROM N.A.  
RP SUBSUB-83590;  
RX MEDLINE-98393590; PubMed-9725850;  
RA Delaire C., Spruyt N., Delmarre C., Gallut C., Barriel V.,

RA Janvier P., Jaudet V., Gachelin G.;  
 RT "The complete nucleotide sequence of the mitochondrial DNA of the  
 RT dogfish, *Scyliorhinus canicula*.";  
 RL Genetics 150:331-344(1998).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.  
 CC -1- CATALYTIC ACTIVITY: 4 ferriocytochrome c + O(2) = 4 ferriocytochrome  
 CC c + 2 H(2)O.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y16067; CAA76022.1; -  
 CC HSSP: P08306; IAR1.  
 CC InterPro: IPR001505; COX2.  
 CC InterPro: IPR002429; Cyt\_c-ox\_2.  
 CC Pfam: PF02790; COX2\_TM; 1.  
 CC Pfam: PF02790; COX2; 1.  
 CC PRINTS: PR01166; CYCOXIDASEII.  
 CC ProDom: PD000131; COX2; 1.  
 CC PROSITE: PS00078; COX2; 1.  
 CC KMW Oxioreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 CC Electron transport; Respiratory chain.  
 CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT TRANSMEM 27 48 PROBABLE.  
 CC FT DOMAIN 49 62 MITOCHONDRIAL MATRIX (POTENTIAL).  
 CC FT TRANSMEM 63 82 PROBABLE.  
 CC FT DOMAIN 83 230 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT METAL 161 161 COPPER A (PROBABLE).  
 CC FT METAL 196 196 COPPER A (PROBABLE).  
 CC FT METAL 200 200 COPPER A (PROBABLE).  
 CC FT METAL 204 204 COPPER A (PROBABLE).  
 CC SEQUENCE 230 AA; 26198 MW; 89311D5BD5BF7397 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALGVKM 22  
 DB 167 ALGVKM 172

RESULT 82  
 COX2\_SQUAC STANDARD; PRT; 230 AA.  
 AC Q92Z51;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
 GN COXII OR COIT.  
 OS Squalus acanthias (Spiny dogfish).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Squalae; Squaloidei; Squalidae; Squalus.  
 OX NCBI\_TaxID=7797;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=99091711; PubMed=9873084;  
 RA Rasmussen A.S., Arnason U.;

RT "Phylogenetic studies of complete mitochondrial DNA molecules place  
 RT cartilaginous fishes within the tree of bony fishes.";  
 RL J. Mol. Evol. 48:118-123(1999).  
 CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME  
 CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA  
 CC HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND  
 CC CU(B). SUBUNIT II BINDS CU(A) AND CYTOCHROME C.  
 CC -1- CATALYTIC ACTIVITY: 4 ferriocytochrome c + O(2) = 4 ferriocytochrome  
 CC c + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A AND HEME GROUP.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial.  
 CC inner membrane.  
 CC -1- SIMILARITY: TO OTHER MITOCHONDRIAL OR BACTERIAL COX2 SUBUNITS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y18134; CAA77052.1; -  
 CC HSSP: P08306; IAR1.  
 CC InterPro: IPR001505; COX2.  
 CC InterPro: IPR002429; Cyt\_c-ox\_2.  
 CC Pfam: PF02790; COX2; 1.  
 CC Pfam: PF02790; COX2\_TM; 1.  
 CC PRINTS: PR01166; CYCOXIDASEII.  
 CC ProDom: PD000131; COX2; 1.  
 CC PROSITE: PS00078; COX2; 1.  
 CC KMW Oxioreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 CC Electron transport; Respiratory chain.  
 CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT TRANSMEM 27 48 PROBABLE.  
 CC FT DOMAIN 49 62 MITOCHONDRIAL MATRIX (POTENTIAL).  
 CC FT TRANSMEM 63 82 PROBABLE.  
 CC FT DOMAIN 83 230 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 CC FT METAL 161 161 COPPER A (PROBABLE).  
 CC FT METAL 196 196 COPPER A (PROBABLE).  
 CC FT METAL 200 200 COPPER A (PROBABLE).  
 CC FT METAL 204 204 COPPER A (PROBABLE).  
 CC SEQUENCE 230 AA; 26246 MW; AC04C362577055A4 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALGVKM 22  
 DB 167 ALGVKM 172

RESULT 83  
 Y394\_PSEAE STANDARD; PRT; 230 AA.  
 AC P24562; P25255;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein PA0394.  
 GN PA0394.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=91285433; PubMed=1676385;  
 RA Whitchurch C.B., Hobbs M., Livingston S.P., Krishnapillai V.,  
 RA Matlick J.S.;  
 RT "Characterisation of a Pseudomonas aeruginosa twitching motility gene

RT and evidence for a specialised protein export system widespread in  
RT eubacteria.";  
RL Gene 101:33-44(1991).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PAOI;  
RX MEDLINE-20437337; PubMed-10994043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.U., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olsson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
RN (3)  
RP SEQUENCE OF 202-230 FROM N.A.  
RC STRAIN-ATCC 15692 / PAOI;  
RX MEDLINE-90185238; PubMed-2107123;  
RA Savioz A., Jeenes D.J., Kocher H.P., Haas D.,  
RT "Comparison of Proc and other housekeeping genes of Pseudomonas  
RT aeruginosa with their counterparts in Escherichia coli.";  
RL Gene 86:107-111(1990).  
CC -1- SIMILARITY: BELONGS TO THE UPE0001 FAMILY.  
CC -1- SIMILARITY: BELONGS TO THE UPE0001 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M55524; AAA25959.1;  
CC EMBL: AE004476; AAG03783.1;  
CC EMBL: M33557; -; NOT\_ANNOTATED; CDS.  
CC PIR: J00060; J00060.  
CC PIR: P00052; P00052.  
CC HSSP: P38197; 1B54.  
CC InterPro: IPR001608; UPE0001.  
CC Pfam: PF01168; UPE0001.1.  
CC PROSITE: PS01211; UPE0001.1.  
CC DR Hypothetical protein; Complete proteome.  
CC KM Hypothetical protein; 24505 MW; 594AE9A1E706768 CRC64;  
CC SQ SEQUENCE 230 AA; 24505 MW; 594AE9A1E706768 CRC64;  
Query Match 5.4%; Score 6; DB 1; Length 230;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 41 ELULD 46  
DB 182 ELULD 187  
RESULT 84  
COX2\_LATCH STANDARD; PRT; 231 AA.  
AC 003846;  
DC 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
GN COXII.  
OS Lactimeria chalumnae (Lactimeria) (Coelacanth).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
OC Coelacanthiformes; Coelacanthidae; Lactimeria.  
OX NCBI\_TaxID=7897;  
RN (1)  
RP SEQUENCE FROM N.A.

RA Zardoya R., Meyer A.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.  
CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
CC c + 2 H(2)O.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
CC inner membrane.  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U82228; AAC60321.1;  
CC HSSP: P08306; IAR1.  
CC InterPro: IPR001505; COX2.  
CC InterPro: IPR002429; Cyt\_c-ox\_2.  
CC Pfam: PF00116; COX2; 1.  
CC DR Pfam: PF02790; COX2\_TM; 1.  
CC DR PRINTS: PR01166; CYCOXIDASEII.  
CC DR PRODOM: PD000131; COX2; 1.  
CC DR PROSITE: PS00078; COX2; 1.  
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
CC KX Electron transport; Respiratory chain.  
CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
FT TRANSMEM 27 48 POTENTIAL.  
FT DOMAIN 49 62 MITOCHONDRIAL MATRIX (POTENTIAL).  
FT TRANSMEM 63 82 POTENTIAL.  
FT DOMAIN 83 231 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
FT METAL 161 161 COPPER A (PROBABLE).  
FT METAL 196 196 COPPER A (PROBABLE).  
FT METAL 200 200 COPPER A (PROBABLE).  
FT METAL 204 204 COPPER A (PROBABLE).  
SQ SEQUENCE 231 AA; 26398 MW; 70FEB6BC09C47FAV CRC64;  
Query Match 5.4%; Score 6; DB 1; Length 231;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 17 ALGVKM 22  
DB 167 ALGVKM 172  
RESULT 85  
RP35\_BACTK STANDARD; PRT; 239 AA.  
ID RP35\_BACTK  
AC P26763;  
DC 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE RNA polymerase sigma-35 factor precursor.  
OS Bacillus thuringiensis (subsp. kurstaki).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=29339;  
RN (1)  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-52.  
RC STRAIN-HD-1-DIPEL.  
RA MEDLINE-91267951; PubMed-1904859.  
RA Adams L.F., Brown K.L., Whiteley H.R.;  
RT "Molecular cloning and characterization of two genes encoding sigma  
RT factors that direct transcription from a Bacillus thuringiensis  
RT crystal protein gene promoter.";

RL J. Bacteriol. 173:3846-3854(1991).  
 [2]  
 RN CHARACTERIZATION.  
 RX MEDLINE-91072211; PubMed-1701426;  
 RA Brown K.L., Whiteley H.R.;  
 RT "Isolation of the second *Bacillus thuringiensis* RNA polymerase that  
 RT transcribes from a crystal protein gene promoter.";  
 RL J. Bacteriol. 172:6682-6688(1990).  
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
 CC THEN IS RELEASED. THIS SIGMA FACTOR DIRECTS TRANSCRIPTION OF  
 CC CRYSAL PROTEIN GENES, A SPOULATION-REGULATED EVENT.  
 CC -1- DEVELOPMENTAL STAGE: IS PROBABLY ACTIVATED AT THE TIME OF SEPTUM  
 CC FORMATION SEPARATING THE MOTHER CELL FROM THE NASCENT ENDOSPORE,  
 CC (STAGE II).  
 CC -1- PTM: ONLY BECOMES ACTIVE AFTER THE PROPEPTIDE IS REMOVED (PROBABLY  
 CC BY A SPOIGA ANALOG) (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X56697; CAA40026.1; -  
 DR PIR: B39441; B39441.  
 DR HSSP: P00579; ISIG.  
 DR InterPro: IPR000943; Sigma\_70.  
 DR Pfam: PF00140; sigma70; 1.  
 DR PROSITE: PS00715; SIGMA70\_1; 1.  
 DR PROSITE: PS00716; SIGMA70\_2; 1.  
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;  
 KW DNA-binding; Spoulation.  
 FT PROPEP 1 27  
 FT CHAIN 28 239 RNA POLYMERASE SIGMA-35 FACTOR.  
 FT DOMAIN 86 99 POLYMERASE CORE BINDING (POTENTIAL).  
 FT DNA\_BIND 206 225 H-T-H MOTIF (BY SIMILARITY).  
 FT CONFLICT 45 45 E -> Y (IN REF. 1; AA SEQUENCE).  
 FT SEQUENCE 239 AA; 27636 MW; FF552F918F93FE24 CRC64;  
 SQ  
 Query Match 5.4%; Score 6; DB 1; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 GNEILL 44  
 Db 148 GNEILL 153  
 RESULT 86  
 RPE\_BACSU STANDARD: PRT; 239 AA.  
 AC P06232; P15810; Q59252;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RNA polymerase sigma-E factor precursor (Sigma-29) (P31) (Stage II  
 DE sporulation protein GB).  
 GN SICE OR SPOIGR.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBL\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-85061557; PubMed-6438529;  
 RA Stragier P., Bouvier J., Bonamy C., Szulmajster J.;  
 RT "A developmental gene product of *Bacillus subtilis* homologous to the  
 RT sigma factor of *Escherichia coli*.";

RL Nature 312:376-378(1984).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP Kobayashi Y., Anaguchi H.;  
 RT "Cloning, amplification, and characterization of sporulation genes,  
 RT especially *spoIIG*, of *Bacillus subtilis*.";  
 RL (in) Setlow P. (eds.);  
 RL Molecular biology of microbial differentiation, pp.85-94, American  
 RL Society for Microbiology, Washington DC (1985).  
 RN [3]  
 RP IDENTIFICATION OF MATURE SIGMA FACTOR AND SEQUENCE OF 30-36.  
 RX MEDLINE-87175561; PubMed-3104904;  
 RA Labell T.L., Trempe J.E., Haldenwang W.G.;  
 RT Sporulation-specific sigma factor sigma 29 of *Bacillus subtilis* is  
 RT synthesized from a precursor protein, *p31*.  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:1784-1788(1987).  
 RN [4]  
 RP SEQUENCE OF 1-55 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE-90174995; PubMed-2106671;  
 RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;  
 RT "Nucleotide sequence of the sporulation gene *spoIIG* from *Bacillus*  
 RT *subtilis*.";  
 RL Nucleic Acids Res. 18:657-657(1990).  
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES  
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND  
 CC THEN IS RELEASED. THIS SIGMA FACTOR IS RESPONSIBLE FOR THE  
 CC EXPRESSION OF SPOULATION SPECIFIC GENES.  
 CC -1- PTM: ONLY BECOMES ACTIVE AFTER THE PROPEPTIDE IS REMOVED (BY  
 CC SPOIGA).  
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X01180; CAA25620.1; -  
 DR EMBL: M57606; AAA22794.1; -  
 DR EMBL: M15804; AAA72566.1; ALT\_SEQ.  
 DR EMBL: X17344; CAA35226.1; -  
 DR EMBL: 299112; CAA13406.1; -  
 DR PIR: S08225; S08225.  
 DR PIR: S07337; S07337.  
 DR Subtilist; BG10235; sigE.  
 DR InterPro: IPR000943; Sigma\_70.  
 DR Pfam: PF00140; sigma70; 1.  
 DR PROSITE: PS00715; SIGMA70\_1; 1.  
 DR PROSITE: PS00716; SIGMA70\_2; 1.  
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;  
 KW DNA-binding; Spoulation; Complete proteome.  
 FT PROPEP 1 29  
 FT CHAIN 30 239 RNA POLYMERASE SIGMA-E FACTOR.  
 FT DOMAIN 86 99 POLYMERASE CORE BINDING (POTENTIAL).  
 FT DNA\_BIND 206 225 H-T-H MOTIF (BY SIMILARITY).  
 FT CONFLICT 139 139 N -> I (IN REF. 2).  
 FT CONFLICT 139 139 D -> Y (IN REF. 2).  
 FT SEQUENCE 239 AA; 27652 MW; 6C517956C4B4CC9 CRC64;  
 SQ  
 Query Match 5.4%; Score 6; DB 1; Length 239;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 GNEILL 44  
 Db 148 GNEILL 153  
 RESULT 87

TRY1\_CHICK STANDARD: PRT: 248 AA.

ID TRY1\_CHICK

AC 090627/

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Trypsin I-P1 precursor (EC 3.4.21.4).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

NCBI\_TaxID=9031;

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RX MEDLINE=95251611; PubMed=7733885;

RA Wang K., Gan L., Lee I., Hood L.E.;

RT "Isolation and characterization of the chicken trypsinogen gene family.";

RL Biochem. J. 307:471-479(1995).

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sb.ch/announce/> or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).

CC -----

CC EMBL: U15156; AAA79912.1;

CC HSP: P00760; IAO7.

CC MEROPS: S01.258; Chymotrypsin.

CC InterPro: IPR001314; Trypsin.

CC Pfam: PF00083; Trypsin.

CC PRINTS: PR00722; Chymotrypsin.

CC SMART: SM00020; Tryp-Spc. 1.

CC PROSITE: PS00240; TRYPSIN\_DOM. 1.

CC PROSITE: PS00134; TRYPSIN\_HIS. 1.

CC PROSITE: PS00135; TRYPSIN\_SER. 1.

CC Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;

KW Multigene family.

KW SIGNAL

FT SIGNAL 1 15

FT PROPEP 16 25

FT CHAIN 26 248

FT ACT\_SITE 65 65

FT ACT\_SITE 109 109

FT ACT\_SITE 202 202

FT DISULFID 32 162

FT DISULFID 50 66

FT DISULFID 134 235

FT DISULFID 141 208

FT DISULFID 173 187

FT DISULFID 198 222

FT SITE 196 196

FT SEQUENCE 248 AA; 26069 MW; C4CF58912B3D58 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 248;

Best Local Similarity 100.0%; Pred No. 1.6e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TISSK 90

DB 89 TISSK 94

RESULT 88

ID TRY2\_CHICK STANDARD: PRT: 248 AA.

AC 090628/

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Trypsin I-P38 precursor (EC 3.4.21.4).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

NCBI\_TaxID=9031;

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas;

RX MEDLINE=95251611; PubMed=7733885;

RA Wang K., Gan L., Lee I., Hood L.E.;

RT "Isolation and characterization of the chicken trypsinogen gene family.";

RL Biochem. J. 307:471-479(1995).

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sb.ch/announce/> or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).

CC -----

CC EMBL: U15156; AAA79913.1;

CC HSP: P00760; IAO7.

CC MEROPS: S01.258; Chymotrypsin.

CC InterPro: IPR001314; Trypsin.

CC Pfam: PF00083; Trypsin.

CC PRINTS: PR00722; Chymotrypsin.

CC SMART: SM00020; Tryp-Spc. 1.

CC PROSITE: PS00240; TRYPSIN\_DOM. 1.

CC PROSITE: PS00134; TRYPSIN\_HIS. 1.

CC PROSITE: PS00135; TRYPSIN\_SER. 1.

CC Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal;

KW Multigene family.

KW SIGNAL

FT SIGNAL 1 15

FT PROPEP 16 25

FT CHAIN 26 248

FT ACT\_SITE 65 65

FT ACT\_SITE 109 109

FT ACT\_SITE 202 202

FT DISULFID 32 162

FT DISULFID 50 66

FT DISULFID 134 235

FT DISULFID 141 208

FT DISULFID 173 187

FT DISULFID 198 222

FT SITE 196 196

FT SEQUENCE 248 AA; 26087 MW; 78B79DD6F15F0CE CRC64;

Query Match 5.4%; Score 6; DB 1; Length 248;

Best Local Similarity 100.0%; Pred No. 1.6e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TISSK 90

DB 89 TISSK 94

```

RESULT 89
H1L_DROVI STANDARD; PRT; 250 AA.
AC Q24704;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1.1.
GN H1S1.1 OR H1S1 OR 12306.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7244;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BOCHUM;
RX MEDLINE=20483906; PubMed=11029073;
RA Nagel S., Grossbach U.;
RT "Histone H1 genes and histone gene clusters in the genus drosophila.";
RL J. Mol. Evol. 51:286-296(2000).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=99132191; PubMed=9933406;
RA Schlemmer J.E., Lozovskaya E.R., Strausbaugh L.D.;
RT "Drosophila virilis has atypical kinds and arrangements of histone
RT repeats.";
RL Chromosoma 107:529-539(1998).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; L76558; AAA0975.1; -.
CC DR HSSP; P02259; 1HST.
CC DR PLAYBASE; PBgn0027800; Dvir\H1S1.1.
CC DR InterPro; IPR001386; Linker_histone.
CC DR InterPro; IPR003216; Linkerhist_N.
CC DR Pfam; PF00538; linker_histone; 1.
CC DR ProDom; PD000373; Linkerhist_N; 1.
CC DR SMART; SM00526; H1S; 1.
CC DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
CC SEQUENCE 250 AA; 25739 MW; C1635F6CED323093 CRC64;

```

```

Query Match 5.4%; Score 6; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.66+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 AVANGN 40
DB 96 AVANGN 101

```

```

RESULT 90
ONCM_HUMAN STANDARD; PRT; 252 AA.
AC P13725;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oncostatin M precursor (OSM).
GN OSM.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89384542; PubMed=2779549;
RA Malik N., Kallestad J.C., Gunderson N.L., Austin S.D., Neubauer M.G.,
RA Ochs V., Margardt H., Zarling J.M., Shoyab M., Wei C.M.,
RA Linsley P.S., Rose T.M.;
RT "Molecular cloning, sequence analysis, and functional expression of a
RT novel growth regulator, oncostatin M.";
RL Mol. Cell. Biol. 9:2847-2853(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Bentley D., Scheet P.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 26-51.
RX MEDLINE=87092323; PubMed=3540948;
RA Zarling J.M., Shoyab M., Margardt H., Hanson M.B., Laublin M.N.,
RA Todaro G.J.;
RT "Oncostatin M: a growth regulator produced by differentiated
RT histiocytic lymphoma cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9739-9743(1986).
RN [4]
RP DISULFIDE BONDS.
RX MEDLINE=9122491; PubMed=2026606;
RA Kallestad J.C., Shoyab M., Linsley P.S.;
RT "Disulfide bond assignment and identification of regions required for
RT functional activity of oncostatin M.";
RL J. Biol. Chem. 266:8940-8945(1991).
RN [5]
RP INHIBITION OF M1 MYELOID LEUKEMIC CELLS.
RX MEDLINE=92020908; PubMed=1717982;
RA Rose T.M., Bruce A.G.;
RT "Oncostatin M is a member of a cytokine family that includes
RT leukemia-inhibitory factor, granulocyte colony-stimulating factor,
RT and interleukin 6.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8641-8645(1991).
RN [6]
RP PROCESSING.
RX MEDLINE=90220571; PubMed=2325640;
RA Linsley P.S., Kallestad J., Ochs V., Neubauer M.;
RT "Cleavage of a hydrophilic C-terminal domain increases
RT growth-inhibitory activity of oncostatin M.";
RL Mol. Cell. Biol. 10:1882-1890(1990).
CC -1- FUNCTION: GROWTH REGULATOR. INHIBITS THE PROLIFERATION OF A NUMBER
CC OF TUMOR CELL LINES. IT REGULATES CYTOKINE PRODUCTION, INCLUDING
CC IL-6, G-CSF AND GM-CSF FROM ENDOTHELIAL CELLS.
CC -1- SIMILARITY: BELONGS TO THE LIF / OSM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; M27288; AAA36388.1; -.
CC DR EMBL; M27286; AAA36388.1; JOINED.
CC DR EMBL; M27287; AAA36388.1; JOINED.
CC DR EMBL; AC004264; AAC05173.1; -.
CC DR PIR; A26663; A26663.
CC DR PIR; A32489; A32489.
CC DR MIM; 165095;
CC DR InterPro; IPR001581; LIF_OSM.
CC DR InterPro; IPR003623; Oncostatin.
CC DR Pfam; PF01291; LIF_OSM; 1.
CC DR ProDom; PD023062; Oncostatin; 1.
CC DR SMART; SM00080; LIF_OSM; 1.
CC DR PROSITE; PS00590; LIF_OSM; 1.
CC Growth regulation; Cytokine; Glycoprotein; Signal.

```

FT SIGNAL 1 25  
FT CHAIN 26 234  
FT PROBE 235 252  
FT DISULFID 31 152  
FT DISULFID 74 192  
FT CARBOHYD 100 100  
FT CARBOHYD 217 100  
FT MUTAGEN 74 74  
FT MUTAGEN 192 192  
FT MUTAGEN 201 201  
FT MUTAGEN 209 209  
SEQUENCE 252 AA; 28484 MW; A5BE281175D101B9 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 252;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVALL 13  
DB 13 LVALL 18

RESULT 91  
CBYC\_ALCEU STANDARD; PRT: 254 AA.  
ID CBYC\_ALCEU P70741;  
AC P40119; P70741;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Cbyp protein, chromosomal.  
OS CBYC OR CBYC.  
GN Alcaligenes eutrophus (Ralstonia eutropha).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=510;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H16 / ATCC 17699;  
RX MEDLINE=93054349; PubMed=1429456;  
RA Kusian B., Yoo J.-G., Bednarski R., Bowlen B.;  
RT "The Calvin cycle enzyme pentose-5-phosphate 3-epimerase is encoded within the cfx operon of the chemototroph Alcaligenes eutrophus.";  
RL J. Bacteriol. 174:7337-7344(1992).  
CC -1- SIMILARITY: BELONGS TO THE CBYC/CBZ/GPH/YIEH FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M64173; AAA21961.1; -  
DR PIR; B47019; B47019.  
DR InterPro; IPR001454; Hydrolase.  
DR Pfam; PF00702; Hydrolase; 1.  
SQ SEQUENCE 254 AA; 27034 MW; 1A6DD9566B3A17CD CRC64;

Query Match 5.4%; Score 6; DB 1; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVL 10  
DB 210 GALLVL 215

RESULT 92  
CBYC\_ALCEU STANDARD; PRT: 254 AA.

AC 004541;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Cbyp protein, plasmid.  
GN CBYC OR CBYC.  
OS Alcaligenes eutrophus (Ralstonia eutropha).  
OC Plasmid megaplasmid pHG1.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=510;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H16 / ATCC 17699;  
RX MEDLINE=93054349; PubMed=1429456;  
RA Kusian B., Yoo J.-G., Bednarski R., Bowlen B.;  
RT "The Calvin cycle enzyme pentose-5-phosphate 3-epimerase is encoded within the cfx operon of the chemototroph Alcaligenes eutrophus.";  
RL J. Bacteriol. 174:7337-7344(1992).  
CC -1- SIMILARITY: BELONGS TO THE CBYC/CBZ/GPH/YIEH FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M64173; AAA98230.1; -  
DR PIR; E47019; E47019.  
DR InterPro; IPR001454; Hydrolase.  
DR Pfam; PF00702; Hydrolase; 1.  
SQ SEQUENCE 254 AA; 27063 MW; D61B01C84FCE1A7 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVL 10  
DB 210 GALLVL 215

RESULT 93  
COAT\_TYDVA STANDARD; PRT: 254 AA.  
ID COAT\_TYDVA P31616;  
AC P31616;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Coat protein.  
GN V2.  
OS Tobacco yellow dwarf virus (strain Australia) (TYDVA).  
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.  
OX NCBI\_TaxID=31599;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92188538; PubMed=1546458;  
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;  
RT "The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants.";  
RL Virology 187:633-642(1992).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: M81103; AAA47948.1; -.
DR PIR: B42452; VCCVTY.
DR InterPro: IPR00143; Gemcoat_MSV.
DR pfam: PF00844; Gemini_coat_Ar1_Br1.
DR PRINTS: PR00223; GEMCOATARB1.
DR ProDom: PD001964; Gemcoat_MSV; 1.
KW Coat protein.
SQ SEQUENCE 254 AA; 28583 MW; 019CF58C3E28B12B CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 254;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 QALGVK 21
Db 22 QALGVK 27

RESULT 94
TRP3_PSEAM
ID TRP3_PSEAM STANDARD; PRT; 256 AA.
AC 093267;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Trypsinogen-like protein 3 precursor.
GN TRP3.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphi; Pleuronectiformes; OC Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX PubMed=9852613;
RA Douglas S.E.; Gallant J.W.;
RT "Isolation of cDNAs for trypsinogen from the winter flounder, Pleuronectes americanus."
RL J. Mar. Biotechnol. 6:214-219(1998).
CC -1- FUNCTION: Not known. Has lost all three of the essential catalytic residues and so probably has no enzymatic activity.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: AF012464; AAC32753.1; -.
DR HSSP: P00766; ICHG.
DR InterPro: IPR001254; Ser-Protase_Try.
DR Pfam: PF00089; trypsin_1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
KW Serine protease homolog; Signal.
FT SIGNAL 1 14
FT CHAIN 15 256 TRYPSINOGEN-LIKE PROTEIN 3.
FT DISULFID 23 153 BY SIMILARITY.
FT DISULFID 41 57 BY SIMILARITY.
FT DISULFID 125 226 BY SIMILARITY.
FT DISULFID 132 199 BY SIMILARITY.
FT DISULFID 164 180 BY SIMILARITY.
FT DISULFID 189 213 BY SIMILARITY.

```

```

SQ SEQUENCE 256 AA; 28524 MW; 68BA7D9265595587 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 256;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LTVLAL 12
Db 4 LTVLAL 9

RESULT 95
GRAM_HUMAN
ID GRAM_HUMAN STANDARD; PRT; 257 AA.
AC P51124;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Granzyme M precursor (EC 3.4.21.-) (Met-ASE) (Natural killer cell granular protease) (HU-Met-1) (Met-1 serine protease).
GN GZMM OR MET1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94065173; PubMed=8245461;
RA Smyth M.J.; Savers T.J.; Wiltout T.; Powers J.C.; Tripani J.A.;
RT "Met-ase: cloning and distinct chromosomal location of a serine protease preferentially expressed in human natural killer cells."
RL J. Immunol. 151:6195-6205(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95229142; PubMed=7713495;
RA Pilat D.; Fink T.M.; Obermaier-Skrobanek B.; Zimmer M.; Wekerle H.;
RA "The human Met-ase gene (GZMM): structure, sequence, and close physical linkage to the serine protease gene cluster on 19p13.3."
RL Genomics 24:445-450(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Smyth M.J.;
RT Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLEAVES PEPTIDE SUBSTRATES AFTER METHIONINE, LEUCINE, AND NORLEUCINE.
CC -1- SUBCELLULAR LOCATION: GRANULES OF LARGE GRANULAR LYMPHOCYTES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY. GRANZYME SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: L36936; AAA57257.1; -.
DR EMBL: L36922; AAA57262.1; -.
DR EMBL: L23134; AAA59582.1; -.
DR HSSP: P20160; 1A7S.
DR MEROPS: S01.139; -.
DR MIM: 600311; -.
DR InterPro: IPR001314; Chymotrypsin.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.

```

KW Hydroxylase: Serine protease; Glycoprotein; Signal; Zymogen; T-cell;  
 KW Cytolysis;  
 FT SIGMA 1 23  
 FT PROPEP 24 25  
 FT CHAIN 26 257  
 FT ACT-SITE 66 66  
 FT ACT-SITE 111 111  
 FT ACT-SITE 207 207  
 FT DISULFID 51 51  
 FT DISULFID 145 213  
 FT DISULFID 176 192  
 FT DISULFID 203 230  
 FT CARBOHYD 177 177  
 SO SEQUENCE 257 AA: 27446 MW: 481500.00 (POTENTIAL).  
 N-TERMINAL (GAGNC...) (POTENTIAL).  
 B1E1SCB5089A801 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLAL 12  
 DB 8 LVLAL 13

RESULT 96  
 ID Y757.METJA STANDARD; PRT; 260 AA.  
 AC O58167;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein M07057;  
 GN M07057;  
 OS Methanococcus jannaschii;  
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus  
 NC NCB1\_TaxID-2190;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE-96337999; PubMed-8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., Fitzgerald L.M., Clifton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.J., Overbeek R., Kirkness E.F., Weinstock K.G., Petráň J., Holtz C.A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Paulsen O., Nguyen D., Uitterback T.R., Kelley J.M., Peterson J.D., Sadelir M., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii";  
 RA Science 273:1058-1073 (1996);  
 RL Science 273:1058-1073 (1996);

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: 067521; AAB98749.1; -  
 DR TIGR: M07057; -  
 DR InterPro: IPR003745; DUF166;  
 DR Pfam: PF02593; DUF166; 1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 260 AA: 30434 MW: 40880.00 (POTENTIAL).  
 F0BBDB3A1143P08 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 260;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 VFFA 37  
 DB 207 VFFA 212

RESULT 98  
 ID Y757.METJA STANDARD; PRT; 261 AA.  
 AC O58167;  
 DT 01-NOV-1997 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein H11086.  
 GN Haemophilus influenzae;  
 OS Haemophilus influenzae;  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 NC NCB1\_TaxID-727;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / RW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uitterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C., Rine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd";  
 RA Science 269:496-512 (1995).  
 RL "SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC - SIMILARITY: STRONG, TO E. COLI Y757.  
 CC - SIMILARITY: TO P. PULVEREA HYPOTHETICAL 28.3 KDA PROTEIN IN YCF26-CHL INTERGENIC REGION (AC P51393).  
 CC CHL INTERGENIC REGION (AC P51393).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U32788; AAC22742.1; -  
 DR TIGR: H11086; -  
 DR InterPro: IPR003453; DUF140;  
 DR Pfam: PF02405; DUF140; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33  
 FT TRANSMEM 30 30  
 FT TRANSMEM 70 70  
 FT TRANSMEM 146 108  
 FT TRANSMEM 199 108  
 FT TRANSMEM 239 259  
 FT TRANSMEM 259 259  
 SO SEQUENCE 261 AA: 28015 MW: 48369.52 (POTENTIAL).  
 4BC3695F247A6BF6 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 VFFA 37  
 DB 207 VFFA 212

RESULT 98  
 ID Y757.METJA STANDARD; PRT; 261 AA.  
 AC O58167;  
 DT 01-NOV-1997 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein H11086.  
 GN Haemophilus influenzae;  
 OS Haemophilus influenzae;  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 NC NCB1\_TaxID-727;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / RW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uitterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C., Rine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd";  
 RA Science 269:496-512 (1995).  
 RL "SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC - SIMILARITY: STRONG, TO E. COLI Y757.  
 CC - SIMILARITY: TO P. PULVEREA HYPOTHETICAL 28.3 KDA PROTEIN IN YCF26-CHL INTERGENIC REGION (AC P51393).  
 CC CHL INTERGENIC REGION (AC P51393).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U32788; AAC22742.1; -  
 DR TIGR: H11086; -  
 DR InterPro: IPR003453; DUF140;  
 DR Pfam: PF02405; DUF140; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33  
 FT TRANSMEM 30 30  
 FT TRANSMEM 70 70  
 FT TRANSMEM 146 108  
 FT TRANSMEM 199 108  
 FT TRANSMEM 239 259  
 FT TRANSMEM 259 259  
 SO SEQUENCE 261 AA: 28015 MW: 48369.52 (POTENTIAL).  
 4BC3695F247A6BF6 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 VFFA 37  
 DB 207 VFFA 212

RESULT 98  
 ID Y757.METJA STANDARD; PRT; 261 AA.  
 AC O58167;  
 DT 01-NOV-1997 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein H11086.  
 GN Haemophilus influenzae;  
 OS Haemophilus influenzae;  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus  
 NC NCB1\_TaxID-727;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / RW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uitterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C., Rine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd";  
 RA Science 269:496-512 (1995).  
 RL "SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC - SIMILARITY: STRONG, TO E. COLI Y757.  
 CC - SIMILARITY: TO P. PULVEREA HYPOTHETICAL 28.3 KDA PROTEIN IN YCF26-CHL INTERGENIC REGION (AC P51393).  
 CC CHL INTERGENIC REGION (AC P51393).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U32788; AAC22742.1; -  
 DR TIGR: H11086; -  
 DR InterPro: IPR003453; DUF140;  
 DR Pfam: PF02405; DUF140; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33  
 FT TRANSMEM 30 30  
 FT TRANSMEM 70 70  
 FT TRANSMEM 146 108  
 FT TRANSMEM 199 108  
 FT TRANSMEM 239 259  
 FT TRANSMEM 259 259  
 SO SEQUENCE 261 AA: 28015 MW: 48369.52 (POTENTIAL).  
 4BC3695F247A6BF6 CRC64;

```

ID  BASI_CHICK  STANDARD:  PRT:  272 AA.
AC  P17790;
DT  01-AUG-1990 (Rel. 15, Created)
DT  01-AUG-1990 (Rel. 15, Last sequence update)
DT  01-MAR-2002 (Rel. 41, Last annotation update)
DE  Basigin precursor (Blood-brain barrier HT7 antigen) (Neurothelin)
GN  (5A11 antigen).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX  MEDLINE-90291979; PubMed-2357963;
RA  Seubert H., Lotzspeich F., Risau W.;
RT  "The inducible blood-brain barrier specific molecule HT7 is a novel
RT  immunoglobulin-like cell surface glycoprotein."
RL  EMBO J. 9:2151-2158(1990).
RN  [2]
RP  PARTIAL SEQUENCE.
RC  STRAIN-WHITE LEDGORN; TISSUE-Retina;
RX  MEDLINE-96128366; PubMed-8575462;
RA  Schlossehauser B., Bauch H., Frank R.;
RT  "Neurothelin: amino acid sequence, cell surface dynamics and actin
RT  colocalization."
RL  Eur. J. Cell Biol. 68:159-166(1995).
CC  -1- FUNCTION: PROBABLE RECEPTOR INVOLVED IN CELL SURFACE RECOGNITION
CC  AT THE BLOOD-BRAIN BARRIER. INVOLVED IN NEURONAL-GLIAL
CC  INTERACTIONS IN RETINAL DEVELOPMENT.
CC  -1- SUBUNIT: HOMODIGOMER.
CC  -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -1- TISSUE SPECIFICITY: ENDOTHELIAL CELLS OF THE CENTRAL NERVOUS
CC  SYSTEM.
CC  -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC  -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC  -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: X52751; CAA36962.1; -.
DR  PIR: S10147; RWCNH7.
DR  InterPro: IPR003599; Ig.
DR  InterPro: IPR003006; Ig_MHC.
DR  Pfam: PF00047; Ig_1.
DR  SMART: SM00409; Ig_1.
KW  Immunoglobulin domain; Transmembrane; Glycoprotein; Signal; Antigen.
FT  SIGNAL 1 26
FT  CHAIN 1 272
FT  DOMAIN 27 210
FT  TRANSMEM 211 234
FT  DOMAIN 235 272
FT  DOMAIN 39 102
FT  DOMAIN 127 197
FT  DISULFID 46 95
FT  DISULFID 134 190
FT  CARBOHYD 47 47
FT  CARBOHYD 106 106
FT  CARBOHYD 164 164
FT  CARBOHYD 170 170
FT  CARBOHYD 191 191
FT  SEQUENCE 272 AA; 29202 MW; 2E8D4CCBCC6A6F2 CRC64;

Query Match 5.4%; Score 6; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 VITALY 14
DB 11 VITALY 16

RESULT 99
TRPA_PORPU
ID TRPA_PORPU STANDARD: PRT: 273 AA.
AC P51382;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Tryptophan synthase alpha chain (EC 4.2.1.20).
GN TRPA.
OS Porphyra purpurea.
OG Chlorophyta.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AVONPORT;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome."
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLYLGLYCEROL PHOSPHATE TO INDOLYL AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC -1- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U38804; AAC08268.1; -.
DR HSSP: P00929; 2MSY.
DR Mendel: I0362; PORPU:trpa.1.
DR InterPro: IPR003009; FMN_enzyme.
DR InterPro: IPR002028; TRP_synthase_alpha.
DR Pfam: PF00290; trp_synta_1.
DR PRODOM: PD001535; TRP_synthase_alpha; 1.
DR PROSITE: PS00167; TRP_SYNTHASE_ALPHA; 1.
KW Tryptophan biosynthesis; Lyase; Chloroplast.
FT SEQUENCE 273 AA; 29835 MW; 2F6C6A072DA8A9AB CRC64;

Query Match 5.4%; Score 6; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 83 MTTISS 88
DB 11 MTTISS 16

RESULT 100
TRYP_PIG
ID TRYP_PIG STANDARD: PRT: 275 AA.
AC Q9N2D1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Trypsin precursor (EC 3.4.21.59).  
GN MCT7.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC Tissue-Lung;  
RA Chen Y., Shiota M., Ohuchi M., Towatari T., Tashiro J., Murakami M.,  
RT Yano M., Yang B., Kido H.;  
RT "A novel trypsin-type protease from pig lungs, triggering infection by  
RT pneumotropic viruses: purification and characterization";  
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADATION  
CC RESPONSE OF THIS CELL TYPE.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-, Lys-I-, but  
CC with more restricted specificity than trypsin.  
CC -1- SUBUNIT: HOMOTETRAMER (By similarity).  
CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
CC MAST CELL ACTIVATION.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AB038652; BAA93614.1; -  
DR MEROPS; S01.143; -  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00089; Trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRY-SPC; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen.  
FT SIGNAL 1 20  
FT PROPEP 21 30  
FT CHAIN 31 275  
FT ACT\_SITE 74  
FT ACT\_SITE 121 121  
FT ACT\_SITE 224 224  
FT DISULFID 39 75  
FT DISULFID 155 230  
FT DISULFID 168 211  
FT DISULFID 220 248  
FT CARBOHYD 132 132  
FT CARBOHYD 223 233  
SO SEQUENCE 275 AA; 30439 MW; ACC582647FCCB973 CRC64;

Query Match 5.48; Score 6; DB 1; Length 275;  
Best Local Similarity 100.08; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLVAL 12  
DB 4 LVLVAL 9

Search completed: July 9, 2002, 16:20:05  
Job time: 115 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2002, 16:18:10 ; Search time 24.46 seconds  
(without alignments)  
785.055 Million cell updates/sec

Title: US-09-662-784-6  
Perfect score: 111  
Sequence: 1 DTMGALLVALVTQALGV.....CMGEAVQNTVEDLKNTLGR 111

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 segs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeoprotein:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	7.2	119	13	Q9YGD0
2	8	7.2	138	13	Q9PSW3
3	8	7.2	138	13	Q9PSW2
4	8	7.2	181	2	Q87052
5	8	7.2	266	2	Q9KW39
6	8	7.2	318	11	Q99NE4
7	8	7.2	322	4	Q9Y577
8	8	7.2	322	4	Q9UNF4
9	8	7.2	390	2	Q9AEX7
10	8	7.2	401	16	Q9HNE4
11	8	7.2	438	11	Q92103
12	8	7.2	438	11	Q92103
13	7	6.3	61	5	Q61042
14	7	6.3	64	5	Q61042
15	7	6.3	73	5	Q9XU39
16	7	6.3	74	5	Q61023

17	7	6.3	77	16	Q9KG69	Q9KG69 bacillus ha
18	7	6.3	78	5	Q26881	Q26881 trypanosoma
19	7	6.3	87	2	Q03539	Q03539 escherichia
20	7	6.3	88	5	Q61047	Q61047 trypanosoma
21	7	6.3	94	5	Q61030	Q61030 trypanosoma
22	7	6.3	102	5	Q61033	Q61033 trypanosoma
23	7	6.3	103	16	Q9R8B9	Q9R8B9 trypanosoma
24	7	6.3	107	5	Q61050	Q61050 trypanosoma
25	7	6.3	112	5	Q9VLE8	Q9VLE8 dirosophila
26	7	6.3	113	5	Q61052	Q61052 trypanosoma
27	7	6.3	115	2	Q9RD28	Q9RD28 streptomyces
28	7	6.3	115	5	Q61046	Q61046 trypanosoma
29	7	6.3	119	5	Q61034	Q61034 trypanosoma
30	7	6.3	122	5	Q15774	Q15774 trypanosoma
31	7	6.3	123	5	Q15773	Q15773 trypanosoma
32	7	6.3	123	5	Q61027	Q61027 trypanosoma
33	7	6.3	123	5	P90602	P90602 trypanosoma
34	7	6.3	125	5	Q61025	Q61025 trypanosoma
35	7	6.3	125	5	Q962W4	Q962W4 trypanosoma
36	7	6.3	126	5	Q61021	Q61021 trypanosoma
37	7	6.3	126	5	Q61056	Q61056 trypanosoma
38	7	6.3	127	5	Q61035	Q61035 trypanosoma
39	7	6.3	128	5	P90603	P90603 trypanosoma
40	7	6.3	133	5	Q61022	Q61022 trypanosoma
41	7	6.3	136	5	Q15775	Q15775 trypanosoma
42	7	6.3	139	5	Q61037	Q61037 trypanosoma
43	7	6.3	139	5	P90601	P90601 trypanosoma
44	7	6.3	139	13	P79799	P79799 micrurus co
45	7	6.3	140	5	Q962W5	Q962W5 trypanosoma

## ALIGNMENTS

RESULT 1  
ID Q9YGD0 PRELIMINARY; PRT; 119 AA.  
AC Q9YGD0;  
DT 01-MAY-1999 (TREMUREL. 10, Created)  
DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)  
DE 01-MAY-1999 (TREMUREL. 10, Last annotation update)  
DE LAMPRLN L-0.9-10.  
GN L-0.9.  
OS Petromyzon marinus (Sea lamprey).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Hyperoartia;  
OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
OX NCBI\_TaxID=7757;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93123269; PubMed=7678258;  
RA Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H., Keeley F.W.;  
RT "Characterization of lampirin, an unusual matrix protein from lamprey cartilage. Implications for evolutionary structure, and assembly of elastin and other fibrillar proteins."  
RT J. Biol. Chem. 268:1440-1447(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Robson P., Youson J.H., Keeley F.W.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF097759; AAC97500.1; -;  
DR EMBL: AF097758; AAC97500.1; JOINED.  
DR EMBL: AF097757; AAC97498.1; -;  
DR EMBL: AF097756; AAC97498.1; JOINED.  
SQ SEQUENCE 119 AA; 11145 MW; BB608EE263DF03E0 CRC64;

Query Match 7.2%; Score 8; DB 13; Length 119;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 6 ALLVAL 13  
|||||||

Db 7 ALLVIAL 14

RESULT 2  
 Q9PSW3 PRELIMINARY; PRT; 138 AA.  
 AC Q9PSW3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE LAMPRIIN L-0.8-12.  
 GN L-0.8.  
 OS Petromyzon marinus (Sea lamprey).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Hypercartia;  
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
 OX NCBI\_TaxID=7757;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93123269; PubMed=7678258;  
 RA Robson P., Wright G.M., Stelz E., Maiti A., Rawat M., Youson J.H.,  
 Keeley F.W.;  
 RT "Characterization of lamprin, an unusual matrix protein from lamprey  
 cartilage. Implications for evolution, structure, and assembly of  
 elastin and other fibrillar proteins.";  
 RL J. Biol. Chem. 268:1440-1447(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20523976; PubMed=11070061;  
 RA Robson P., Wright G.M., Youson J.H., Keeley F.W.;  
 RT "The structure and organization of lamprin genes: multiple-copy genes  
 with alternative splicing and convergent evolution with insect  
 structural proteins.";  
 RL Mol. Biol. Evol. 17:1739-1752(2000).  
 DR EMBL: AF097757; AAC97499.1; -;  
 DR EMBL: AF097756; AAC97499.1; JOINED.  
 SO SEQUENCE 138 AA; 13094 MW; AB2F35D0D25D7580 CRC64;

Query Match 7.2%; Score 8; DB 13; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 13  
 Db 7 ALLVIAL 14

RESULT 3  
 Q9PSW2 PRELIMINARY; PRT; 138 AA.  
 AC Q9PSW2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE LAMPRIIN L-0.9-12.  
 GN L-0.9.  
 OS Petromyzon marinus (Sea lamprey).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Hypercartia;  
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
 OX NCBI\_TaxID=7757;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93123269; PubMed=7678258;  
 RA Robson P., Wright G.M., Stelz E., Maiti A., Rawat M., Youson J.H.,  
 Keeley F.W.;  
 RT "Characterization of lamprin, an unusual matrix protein from lamprey  
 cartilage. Implications for evolution, structure, and assembly of  
 elastin and other fibrillar proteins.";  
 RL J. Biol. Chem. 268:1440-1447(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20523976; PubMed=11070061;  
 RA Robson P., Wright G.M., Youson J.H., Keeley F.W.;

RT "The structure and organization of lamprin genes: multiple-copy genes  
 with alternative splicing and convergent evolution with insect  
 structural proteins.";  
 RL Mol. Biol. Evol. 17:1739-1752(2000).  
 DR EMBL: AF097759; AAC97501.1; -;  
 DR EMBL: AF097758; AAC97501.1; JOINED.  
 SO SEQUENCE 138 AA; 13144 MW; AB2F35D0C74924D0 CRC64;

Query Match 7.2%; Score 8; DB 13; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 13  
 Db 7 ALLVIAL 14

RESULT 4  
 Q87052 PRELIMINARY; PRT; 181 AA.  
 AC Q87052;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 248F PROTEIN (FRAGMENT).  
 GN 248F.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=217561.  
 RA Fallarino A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A2231108; CA13150.1; -;  
 DR InterPro: IPR003764; NaGA.  
 DR Pfam: PF02612; NaGA.1.  
 FT NON\_TER 1  
 FT NON\_TER 181  
 FT NON\_TER 181  
 SO SEQUENCE 181 AA; 19476 MW; 8527718DAC9DD199 CRC64;

Query Match 7.2%; Score 8; DB 2; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 EAVONTVE 102  
 Db 150 EAVONTVE 157

RESULT 5  
 Q9KW39 PRELIMINARY; PRT; 266 AA.  
 AC Q9KW39;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE COMPONENT OF TYPE IV SECRETION SYSTEM.  
 GN VIB9.  
 OS Wolbachia sp. wkuexy.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Wolbachiae; Wolbachia.  
 OX NCBI\_TaxID=130485;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=WKUEYO;  
 RC MEDLINE=20507828; PubMed=11053403;  
 RA Masui S., Sasaki T., Ishikawa H.;  
 RT "Genes for the Type IV Secretion System in an Intracellular Symbiont,  
 Wolbachia, a Causative Agent of Various Sexual Alterations in  
 Arthropods.";

RL J. Bacteriol. 182:6529-6531(2000).  
 DR EMBL: A8045235; BAA97440.1; -  
 SO SEQUENCE 266 AA; 3112 MW; 23CA2E77A756B5E2 CRC64;

Query Match 7.2%; Score 8; DB 2; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LVLVALL 14  
 Db 7 LVLVALL 14

RESULT 6  
 O99NE4 PRELIMINARY; PRT; 318 AA.  
 AC O99NE4;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYALURONAN RECEPTOR PRECURSOR.  
 GN XKDI OR LYVE-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=DIGESTIVE TRACT;  
 RX MEDLINE=99156989; PubMed=10037799;  
 RA Banerji S.;  
 RT "LYVE-1, a new homologue of the CD44 glycoprotein is a lymph-specific  
 RT receptor for hyaluronan.";  
 RL J. Cell Biol. 144:789-801(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=DIGESTIVE TRACT;  
 RX MEDLINE=21276443; PubMed=11278811;  
 RA Prevost R., Banerji S., Ferguson D.J.P., Clasper S., Jackson D.G.;  
 RT "Mouse LYVE-1 is an endocytic receptor for hyaluronan in lymphatic  
 RT endothelium.";  
 RL J. Biol. Chem. 276:19420-19430(2001).  
 DR EMBL: AFJ11501; CAC33082.1; -  
 DR HSSP: P98066; ITSG.  
 DR MGD: MGI:2136348; Xlkd1.  
 DR InterPro: IPR000538; Link.  
 DR InterPro: IPR001190; SRCR.  
 DR Pfam: PF00193; Xlink; 1.  
 DR ProDom: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 DR SMART: SM00202; SR; 1.  
 DR PROSITE: PS01241; LINK; UNKNOWN\_1.  
 DR Signal; Receptor.  
 KW SIGNAL; 23  
 FT CHAIN 1 24 318 LYVE-1.  
 SO SEQUENCE 318 AA; 34641 MW; 1248974A16113330 CRC64;

Query Match 7.2%; Score 8; DB 11; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVALL 13  
 Db 235 ALLVALL 242

RESULT 7  
 O9Y5Y7 PRELIMINARY; PRT; 322 AA.  
 ID O9Y5Y7  
 AC O9Y5Y7;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE LYMPHATIC ENDOTHELIAL-SPECIFIC HYALURONAN RECEPTOR LYVE-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99156989; PubMed=10037799;  
 RA Banerji S., Ni J., Wang S.X., Clasper S., Su J., Tammi R., Jones M.,  
 RA Jackson D.G.;  
 RT "LYVE-1, a new homologue of the CD44 glycoprotein, is a lymph-specific  
 RT receptor for hyaluronan.";  
 RL J. Cell Biol. 144:789-801(1999).  
 DR EMBL: AF118108; AAD42764.1; -  
 DR HSSP: P98066; ITSG.  
 DR InterPro: IPR000538; Link.  
 DR Pfam: PF00193; Xlink; 1.  
 DR ProDom: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 KW RECEPTOR.  
 SO SEQUENCE 322 AA; 35238 MW; 0EBEA56729CEFF7 CRC64;

Query Match 7.2%; Score 8; DB 4; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVALL 13  
 Db 239 ALLVALL 246

RESULT 8  
 O9UNF4 PRELIMINARY; PRT; 322 AA.  
 ID O9UNF4;  
 AC O9UNF4;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE HYALURONIC ACID RECEPTOR.  
 GN HAR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Winkelman J.C., Basu S., Ozdemir E., Blough R.I.;  
 RT "HAR: a novel homolog of CD44 and putative hyaluronic acid receptor  
 RT encoded by a gene on human chromosome 11p15.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF127670; AAD49220.2; -  
 DR HSSP: P98066; ITSG.  
 DR InterPro: IPR000538; Link.  
 DR Pfam: PF00193; Xlink; 1.  
 DR PRINTS: PR01265; LINKMODULE.  
 DR ProDom: PD000918; Link; 1.  
 DR SMART: SM00445; Link; 1.  
 KW RECEPTOR.  
 SO SEQUENCE 322 AA; 35213 MW; 8B4D6D623F52D559 CRC64;

Query Match 7.2%; Score 8; DB 4; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVALL 13  
 Db 239 ALLVALL 246

## RESULT 9

09AEX7 PRELIMINARY; PRT; 390 AA.  
ID 09AEX7  
AC 09AEX7  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VARIABLE SURFACE PROTEIN VSPF.  
GN VSPF.  
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).  
OC Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.  
NCBI\_TaxID=159;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCaman M., Gabe J.;  
RT "Serpulina hyodysenteriae (B204) contains eight gene copies related to a 39 kDa surface protein."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY027775; AAK14801.1;  
SQ SEQUENCE 390 AA; 42836 MW; 9B20084CF61A2B9E CRC64;

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GEAVONTV 101  
DB 247 GEAVONTV 254  
|||||

## RESULT 10

09AEX5 PRELIMINARY; PRT; 390 AA.  
ID 09AEX5  
AC 09AEX5  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VARIABLE SURFACE PROTEIN VSPF.  
GN VSPF.  
OS Treponema hyodysenteriae (Serpulina hyodysenteriae).  
OC Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.  
NCBI\_TaxID=159;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McCaman M., Gabe J.;  
RT "Serpulina hyodysenteriae (B204) contains eight gene copies related to a 39 kDa surface protein."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY027775; AAK14801.1;  
SQ SEQUENCE 390 AA; 42816 MW; C3B63F5C9BC36E08 CRC64;

## Query Match

Best Local Similarity 7.2%; Score 8; DB 2; Length 390;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GEAVONTV 101  
DB 245 GEAVONTV 252  
|||||

## RESULT 11

09HT26 PRELIMINARY; PRT; 401 AA.  
ID 09HT26  
AC 09HT26  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE PROBABLE MFS TRANSPORTER.  
GN PA5548.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-ATCC 15692 / PA01.  
RX DDBLINK=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.T., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen."  
RL Nature 406:959-964(2000).  
DR EMBL: AE004967; AAC08933.1;  
DR InterPro: IPR001958; TCR\_TetA.  
DR PRINTS: PR01035; TCRTEA.  
DR PRINTS: PR01036; TCRTEB.  
KW Complete proteome.  
SQ SEQUENCE 401 AA; 42247 MW; D8D9E8ADBD509E CRC64;

## Query Match

Best Local Similarity 7.2%; Score 8; DB 16; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LTVLALLV 14  
DB 281 LTVLALLV 288  
|||||

## RESULT 12

092103 PRELIMINARY; PRT; 438 AA.  
ID 092103  
AC 092103  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SIMILAR TO BETA-1,4 MANNOSYLTRANSFERASE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC011281; AAH1281.1;  
RW Transferase; Glycosyltransferase.  
SQ SEQUENCE 438 AA; 49698 MW; 06AA82FE1B95908 CRC64;

## Query Match

Best Local Similarity 7.2%; Score 8; DB 11; Length 438;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALTVLALL 13  
DB 7 ALTVLALL 14  
|||||

## RESULT 13

061029 PRELIMINARY; PRT; 61 AA.  
ID 061029  
AC 061029  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
GN EMUCE-21E22BIS.  
OS Trypanosoma cruzi.



OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RT genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL: AF036418; AAC14228.1; -.  
FT NON-TER 61 61  
SQ SEQUENCE 61 AA: 6192 MW: 95ABC7D84C05B587 CRC64;  
  
Query Match 6.3%; Score 7; DB 5; Length 61;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 ALLVTAL 12  
DB 10 ALLVTAL 16  
  
RESULT 14  
061042 PRELIMINARY; PRT; 64 AA.  
AC 061042;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
GN EMUCE-13.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RT genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL: AF036418; AAC14228.1; -.  
FT NON-TER 64 64  
SQ SEQUENCE 64 AA: 6673 MW: A2B55B7D05608FF CRC64;  
  
Query Match 6.3%; Score 7; DB 5; Length 64;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 ALLVTAL 12  
DB 14 ALLVTAL 20  
  
RESULT 15  
09XU39 PRELIMINARY; PRT; 73 AA.  
AC 09XU39;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE F14F8.8 PROTEIN.  
GN F14F8.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Lloyd C.R.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL: Z92782; CAB07183.1; -.  
SQ SEQUENCE 73 AA: 7210 MW: 2F5F985045CB141D CRC64;  
  
Query Match 6.3%; Score 7; DB 5; Length 73;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 LVLVALL 13  
DB 5 LVLVALL 11  
  
RESULT 16  
061023 PRELIMINARY; PRT; 74 AA.  
AC 061023;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
GN EMUCE-4.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RT genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL: AF036411; AAC14222.1; -.  
FT NON-TER 74 74  
SQ SEQUENCE 74 AA: 7743 MW: 73ACC37663E21401 CRC64;  
  
Query Match 6.3%; Score 7; DB 5; Length 74;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 ALLVTAL 12  
DB 14 ALLVTAL 20  
  
RESULT 17  
09KG69 PRELIMINARY; PRT; 77 AA.  
AC 09KG69;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE BH0244 PROTEIN.  
GN BH0244.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kihara S.,  
 RA Horikoshi K.; 267:20471-20480(1992).  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AF001507; BAB03963.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 77 AA; 9511 MW; F9BEC3529F5CE320 CRC64;

Query Match 6.3%; Score 7; DB 16; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 41 ELLLDLS 47  
 |||||  
 DB 24 ELLLDLS 30

RESULT 18  
 ID 026881 PRELIMINARY; PRT; 78 AA.  
 AC 026881;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE MUCIN (FRAGMENT).  
 GN MUC-CA-1.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CA1/72;  
 RX MEDLINE=96025798; PubMed=7592617;  
 RA Di Nola J.M., Sanchez D.O., Frasch A.C.;  
 RT "The protozoan Trypanosoma cruzi has a family of genes resembling the  
 RT mucin genes of mammalian cells.";  
 RL J. Biol. Chem. 270:24146-24149(1995).  
 DR EMBL; U32572; AAC6905.1; -  
 FT NON-TER 1 78  
 SQ SEQUENCE 78 AA; 8056 MW; 4B869C9014FB62CF CRC64;

Query Match 6.3%; Score 7; DB 5; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 ALLVIAL 12  
 |||||  
 DB 7 ALLVIAL 13

RESULT 19  
 ID 003539 PRELIMINARY; PRT; 87 AA.  
 AC 003539;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE TRBO PROTEIN.  
 GN TRBO.  
 OS Escherichia coli.  
 OC Bacteriia; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93015931; PubMed=1400366;  
 RA Lessl M., Balzer D., Pansegrau W., Lanka E.;

RT "Sequence similarities between the RP4 Trz2 and the T1 VirB region  
 RT strongly support the conjugation model for T-DNA transfer.";  
 RL J. Biol. Chem. 267:20471-20480(1992).  
 DR EMBL; M93696; AAA26441.1; -  
 KW Plasmid.  
 SQ SEQUENCE 87 AA; 9440 MW; 80424E9A493D8C5 CRC64;

Query Match 6.3%; Score 7; DB 2; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 GALLVIA 11  
 |||||  
 DB 62 GALLVIA 68

RESULT 20  
 ID 061047 PRELIMINARY; PRT; 88 AA.  
 AC 061047;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 GN EMUC-8.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions.";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DR EMBL; AF036451; AAC14248.1; -  
 DR InterPro; IPR000458; Tryp\_mucin.  
 DR Pfam; PF01456; Tryp\_mucin.1.  
 SQ SEQUENCE 88 AA; 9092 MW; A335AB30729C7C5 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 ALLVIAL 12  
 |||||  
 DB 14 ALLVIAL 20

RESULT 21  
 ID 061030 PRELIMINARY; PRT; 94 AA.  
 AC 061030;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 GN EMUC-6.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions.";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DR EMBL; AF036419; AAC14229.1; -

FT NON\_TER 94 94  
SQ SEQUENCE 94 AA; 9840 MW; 3A79EED1E7512422 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 5; Length 94;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
Db 14 ALLVIAL 20

RESULT 22  
061033 PRELIMINARY; PRT; 102 AA.

AC 061033;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
GN EMUC-11.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxId=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA MEDLINE=98225151; PubMed=9556557;  
DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RL genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL; AF036422; AAC14232.1;  
DR InterPro; IPR000458; TRYP\_mucin.  
DR Pfam; PF01456; TRYP\_mucin; 1.  
FT NON\_TER 102  
SQ SEQUENCE 102 AA; 10605 MW; E55212A8D1297E5A CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 5; Length 102;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
Db 14 ALLVIAL 20

RESULT 23  
09RRB9 PRELIMINARY; PRT; 103 AA.

AC 09RRB9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE HYPOTHEICAL 10.5 KDA PROTEIN.  
GN DR2573.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus; Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxId=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI;  
RA MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE002086; AAF12117.1; -.

DR TIGR; DR2573; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 103 AA; 10517 MW; BC814494FF991CCD CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 16; Length 103;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
Db 15 ALLVIAL 21

RESULT 24  
061050 PRELIMINARY; PRT; 107 AA.

AC 061050;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
GN EMUC-9.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxId=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA MEDLINE=98225151; PubMed=9556557;  
DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RL genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL; AF036454; AAC14251.1;  
DR InterPro; IPR000458; TRYP\_mucin.  
DR Pfam; PF01456; TRYP\_mucin; 1.  
FT NON\_TER 107  
SQ SEQUENCE 107 AA; 10986 MW; 26E2947FDEB06D2 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 5; Length 107;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
Db 9 ALLVIAL 15

RESULT 25  
09VLE8 PRELIMINARY; PRT; 112 AA.

AC 09VLE8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
DE CG13105 PROTEIN.  
GN CG13105.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazet R.C., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson G.R., Milos G.L.G.,  
 RA Abell J.F., Appayani A., An H.-J., Andrews-Fiankoch C., Baldwin D.,  
 RA Ballew R.M., Bess A., Bakendale J., Bakker-Elstgen L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernat B., Benhardt D., Bolshakov S.,  
 RA Borkova D., Botchan M.K., Bouck H., Cadieu E., Center A., Chandra I.,  
 RA Butts K.C., Busan D.A., Butler H., Cadenot L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Deng Z., Dempsey L.B., Deyam C.,  
 RA de Pablo B., Decher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup E.R., Downes M., Dugan-Kocha S., Dunn B.C., Dunn P.,  
 RA Dushin K.J., Pyangellista C.C., Feig M.S., Gelbart W.M., Glasser K.,  
 RA Foster C., Gabrielian A.E., Gang N.S., Gellera S., Fleischmann W.,  
 RA Glodok A., Gong P., Correll J.H., Guan P., Harris M.,  
 RA Hostin D., Houston K., Hokland T.J., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Helander J., Mel M.-H., Idegami C.,  
 RA Jajali M., Kalush F., Karpel G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Lin X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Lin X.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Lin X.,  
 RA Mekulov G., Milshina N.Y., Mobarry C., Morris M., Moser A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D., Nadeau J.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Packer J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith R.,  
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtskas R., Tector C., Turner R., Ventier E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu S., Zhu X., Smith H.O.,  
 RA Zheng X.H., Zhong F.N., Zhong G.M., Wetschbach T.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of *Drosophila melanogaster*,"  
 RT Science 287:2185-2195(2000).  
 RL EMBL: AEO03623; AAF52745.1;  
 DR FlyBase: FBgn0032093; CG13105.  
 DR SEQUENCE 112 AA; 10259 MW; A78B1DD9FC493750 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 5 GALLVLA 11  
 DB 10 GALLVLA 16  
 RESULT 26  
 ID 061052 PRELIMINARY; PRT; 113 AA.  
 AC 061052;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 GN EMUCT-11.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 RX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DR EMBL: AF036456; AAC14253.1;  
 DR InterPro: IPR000458; Tryp\_mucin; Pfam: PF01456; Tryp\_mucin; 1.

FT NON-TER 113 113  
 SQ SEQUENCE 113 AA; 11604 MW; 6A25FFC9670267F0 CRC64;  
 Query Match 6.3%; Score 7; DB 5; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 6 ALLVVAL 12  
 DB 9 ALLVVAL 15

RESULT 27  
 ID 09RD28 PRELIMINARY; PRT; 115 AA.  
 AC 09RD28;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOHETICAL 13.2 KDA PROTEIN.  
 GN SCMA-15.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycesaceae; Streptomyces.  
 RX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D.C., Harris D.;  
 RU Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Thomson N.R., Parhill J., Bartell B.G., Rajandream M.A.;  
 RU Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleiser H.M., Denapette D., Eichner A., Cullum J.,  
 RA Kinslin H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome,";  
 RL Mol. Microbiol. 23:77-92(1996).  
 DR EMBL: AL13342; CAB62673.1;  
 DR InterPro: IPR001460; TPR.  
 KM Hypothetical Protein.  
 SQ SEQUENCE 115 AA; 13167 MW; 5D6D78AB32CAAF0 CRC64;

Query Match 6.3%; Score 7; DB 2; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 76 RVLDGLV 82  
 DB 32 RVLDGLV 38  
 RESULT 28  
 ID 061046 PRELIMINARY; PRT; 115 AA.  
 AC 061046;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 GN EMUCT-7.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 RX NCBI\_TaxID=5693;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions.";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DE EMBL: AF036450; AACC14247.1; -;  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 SQ SEQUENCE 115 AA; 11729 MW; 321826F0FDEDF0E CRC64;

Query Match 6.3%; Score 7; DB 5; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

RESULT 29  
 ID 061034 PRELIMINARY; PRT; 119 AA.  
 AC 061034;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 GN EMUC-2409.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions.";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DE EMBL: AF036424; AACC14233.1; -;  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 119  
 SQ SEQUENCE 119 AA; 12072 MW; C86DEAB036D4B38F CRC64;

Query Match 6.3%; Score 7; DB 5; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 10 ALLVIAL 16

RESULT 30  
 ID 015774 PRELIMINARY; PRT; 122 AA.  
 AC 015774;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RT MEDLINE=98324409; PubMed=9662032;

RA Freitas-Junior L.H., Briones M.R., Schenkman S.;  
 RT "Two distinct groups of mucin-like genes are differentially expressed  
 RT in the developmental stages of Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 93:101-114(1998).  
 DE EMBL: AF027872; AAC48350.1; -;  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 122  
 SQ SEQUENCE 122 AA; 12500 MW; 47CDEF9BDA3814FA CRC64;

Query Match 6.3%; Score 7; DB 5; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

RESULT 31  
 ID 015773 PRELIMINARY; PRT; 123 AA.  
 AC 015773;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RX MEDLINE=98324409; PubMed=9662032;  
 RA Freitas-Junior L.H., Briones M.R., Schenkman S.;  
 RT "Two distinct groups of mucin-like genes are differentially expressed  
 RT in the developmental stages of Trypanosoma cruzi.";  
 RL Mol. Biochem. Parasitol. 93:101-114(1998).  
 DE EMBL: AF027871; AAC48349.1; -;  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 123  
 SQ SEQUENCE 123 AA; 12574 MW; 29A53DCB877C1385 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

RESULT 32  
 ID 061027 PRELIMINARY; PRT; 123 AA.  
 AC 061027;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 GN EMUC-10.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions.";

RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL: AF036413; AAC14226.1; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 123 AA; 12742 MW; 0FDD570E5C91308D CRC64;

Query Match 6.3%; Score 7; DB 5; Length 123;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

RESULT 33  
ID P90602 PRELIMINARY; PRT; 123 AA.  
AC P90602.  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUC-TM-1.  
GN MUC-TM-1.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA DI Noia J.M., Pollevick G.D., Xavier M.T., Previalto J.O.,  
RA Mendoca-Previalto L., Sanchez D.O., Frasch A.C.;  
RT "High diversity in mucin genes and mucin molecules in Trypanosoma  
RT cruzi.";  
RL J. Biol. Chem. 271:32078-32083(1996).  
DR EMBL: U61532; AAC47400.1; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 123 AA; 12814 MW; 157656BE5C91308D CRC64;

Query Match 6.3%; Score 7; DB 5; Length 123;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

RESULT 34  
ID 061025 PRELIMINARY; PRT; 125 AA.  
AC 061025;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUCE-9.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA MEDLINE=98225151; PubMed=9556557;  
RA DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RT genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL: AF036413; AAC14224.1; -

DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 125 AA; 12894 MW; 2DF1A14AA29A8604 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 125;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

RESULT 35  
ID 0962M4 PRELIMINARY; PRT; 125 AA.  
AC 0962M4;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN MUC-LOC6.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA DI Noia J.M., Frasch A.C.C.;  
RT "Hypervariable mucin genes from Trypanosoma cruzi.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF398553; AAK94016.1; -  
SQ SEQUENCE 125 AA; 12870 MW; 2188F87FA6C71F07 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 125;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

RESULT 36  
ID 061021 PRELIMINARY; PRT; 126 AA.  
AC 061021;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUCE-2.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA MEDLINE=98225151; PubMed=9556557;  
RA DI Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
RT genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CL-BRENNER;  
RA D'Orso I., DI Noia J.M.;  
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF036409; AAC14220.2; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 126 AA; 13023 MW; F385808D3C768A1 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 126;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

## RESULT 37

061056 PRELIMINARY; PRT; 126 AA.  
ID 061056  
AC 061056;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUC-15.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RA D'Orso I., Di Noia J.M.;  
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF036463; AAC14257.2; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 126 AA; 13049 MW; F399EC78D3C768A1 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 126;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

## RESULT 38

061035 PRELIMINARY; PRT; 127 AA.  
ID 061035  
AC 061035;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUC-3.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL; AF036425; AAC14234.1; -  
DR InterPro: IPR000458; Tryp\_mucin.

DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 127 AA; 12958 MW; 37CAB5052BF3B79B CRC64;

Query Match 6.3%; Score 7; DB 5; Length 127;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 10 ALLVIAL 16

## RESULT 39

P90603 PRELIMINARY; PRT; 128 AA.  
ID P90603  
AC P90603;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MUC.CL-1.  
GN MUC.CL-1.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=97113006; PubMed=8943259;  
RA Di Noia J.M., Pollevick G.D., Xavier M.T., Previanto J.O.,  
RA Mendoca-Previanto L., Sanchez D.O., Frasch A.C.;  
RT "High diversity in mucin genes and mucin molecules in Trypanosoma  
cruzi.";  
RL J. Biol. Chem. 271:32078-32083(1996).  
DR EMBL; U62530; AAC47402.1; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 128 AA; 13207 MW; 30ACB7C3F8E633B4 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 128;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ALLVIAL 12  
| | | | |  
Db 14 ALLVIAL 20

## RESULT 40

061022 PRELIMINARY; PRT; 133 AA.  
ID 061022  
AC 061022;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUC-3.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
genes having hypervariable regions.";  
RL J. Biol. Chem. 273:10843-10850(1998).  
DR EMBL; AF036410; AAC14221.1; -  
DR InterPro: IPR000458; Tryp\_mucin.  
DR Pfam: PF01456; Tryp\_mucin; 1.  
SQ SEQUENCE 133 AA; 13692 MW; 9FC8295E10196D81 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 133;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
|||||  
DB 14 ALLVIAL 20

RESULT 41  
ID 015775 PRELIMINARY; PRT; 136 AA.  
AC 015775;  
DT 01-JAN-1998 (TReMBLrel. 05, Created)  
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN (FRAGMENT).  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Y;  
RX MEDLINE=98324409; PubMed=9662032;  
RA Freitas-Junior L.H., Briones M.R., Schenkman S.;  
RT "Two distinct groups of mucin-like genes are differentially expressed  
in the developmental stages of Trypanosoma cruzi.";  
RT Mol. Biochem. Parasitol. 93:101-114(1998).  
RL EMBL: AF027873; AAC48351.1;  
DR InterPro: IPR000458; TRYP\_mucin.  
DR Pfam: PF01456; TRYP\_mucin; 1.  
FT NON\_TER 136  
SQ SEQUENCE 136 AA; 13666 MW; EEl2C4705586142B CRC64;

Query Match 6.3%; Score 7; DB 5; Length 136;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
|||||  
DB 9 ALLVIAL 15

RESULT 42  
ID 061037 PRELIMINARY; PRT; 139 AA.  
AC 061037;  
DT 01-AUG-1998 (TReMBLrel. 07, Created)  
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE MUCIN-LIKE PROTEIN.  
GN EMUCE-37P20.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CL-BRENNER;  
RX MEDLINE=98225151; PubMed=9556557;  
RA Di Noia J.M., D'Orso I., Ashund L., Sanchez D.O., Frasch A.C.;  
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
genes having hypervariable regions.";  
RT J. Biol. Chem. 273:10843-10850(1998).  
RL EMBL: AF036427; AAC14349.1;  
DR InterPro: IPR000458; TRYP\_mucin.  
DR Pfam: PF01456; TRYP\_mucin; 1.  
SQ SEQUENCE 139 AA; 14311 MW; 9236BB31B8599287 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 139;

Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
|||||  
DB 14 ALLVIAL 20

RESULT 43  
ID P90601 PRELIMINARY; PRT; 139 AA.  
AC P90601;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE MUC Y-1 PROTEIN.  
GN MUC Y-1.  
OS Trypanosoma cruzi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Y;  
RX MEDLINE=97113006; PubMed=8943259;  
RA Di Noia J.M., Polievick G.D., Xavier M.T., Previalto J.O.,  
RA Mendoca-Previalto L., Sanchez D.O., Frasch A.C.;  
RT "High diversity in mucin genes and mucin molecules in Trypanosoma  
cruzi.";  
RT J. Biol. Chem. 271:32078-32083(1996).  
RL EMBL: U59482; AAC47399.1;  
DR InterPro: IPR000458; TRYP\_mucin.  
DR Pfam: PF01456; TRYP\_mucin; 1.  
SQ SEQUENCE 139 AA; 14395 MW; D7DCCEB2F8A26B CRC64;

Query Match 6.3%; Score 7; DB 5; Length 139;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
|||||  
DB 14 ALLVIAL 20

RESULT 44  
ID P79799 PRELIMINARY; PRT; 139 AA.  
AC P79799;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE NATRIURETIC PEPTIDE.  
OS Micrurus corallinus (Brazilian coral snake).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubridae;  
OC Elapidae; Elapinae; Micrurus.  
OX NCBI\_TaxID=54390;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND;  
RA Ho P.L., Soares M.B., Yamane T., Raw I.;  
RT "Reverse biology applied to Micrurus corallinus, a South American  
coral snake.";  
RT J. Toxicol. Toxin. Rev. 14:327-337(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=VENOM GLAND;  
RX MEDLINE=98092299; PubMed=9432002;  
RA Ho P.L., Soares M.B., Maack T., Gimenez I., Puerto G., Furtado M.F.,  
RA Raw I.;  
RT "Cloning of an unusual natriuretic peptide from the South American  
coral snake Micrurus corallinus.";  
RT Eur. J. Biochem. 250:144-149(1997).



DR EMBL: U77596; AAC60341.1; -  
 DR InterPro: IPR002408; Br\_natriurctpep.  
 DR InterPro: IPR000663; Natf\_peptide.  
 DR Pfam: PF00212; ANP; 1.  
 DR PRINTS: PR00710; NATPEPTIDES.  
 DR ProDom: PD006651; Br\_natriurctpep; 1.  
 DR SMART: SM00183; NAT\_PEP; 1.  
 DR PROSITE: PS00263; NATRIURETIC\_PEPTIDE; 1.  
 SQ SEQUENCE 139 AA; 14881 MW; 426287771870E1PB CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 13; Length 139;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLVALL 13  
 |||||  
 DB 13 LVLVALL 19

RESULT 45  
 Q962W5 PRELIMINARY; PRT; 140 AA.  
 AC Q962W5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN MUC-LOC5.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RA Di Noia J.M., Frasch A.C.C.;  
 RT "Hypervariable mucin genes from Trypanosoma cruzi."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF398552; AAK94015.1; -  
 SQ SEQUENCE 140 AA; 14343 MW; 5CC154418F2A58CA CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 5; Length 140;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 DB 14 ALLVIAL 20

RESULT 46  
 O15776 PRELIMINARY; PRT; 143 AA.  
 ID O15776;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RA MEDLINE=98324409; PubMed=9662032;  
 RA Freitas-Junior L.H., Bioness M.R., Schenkman S.;  
 RT "Two distinct groups of mucin-like genes are differentially expressed  
 in the developmental stages of Trypanosoma cruzi."  
 RL Mol. Biochem. Parasitol. 93:101-114(1998).  
 DR EMBL: AF027874; AAC48352.1; -  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 FT NON\_TER 143 143

SQ SEQUENCE 143 AA; 14610 MW; 6AB6E7B7FA85F58B CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 5; Length 143;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 DB 9 ALLVIAL 15

RESULT 47  
 O61043 PRELIMINARY; PRT; 147 AA.  
 ID O61043;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 GN EMUCT-2.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RA MEDLINE=98225151; PubMed=9556557;  
 RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions."  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DR EMBL: AF036441; AAC14243.1; -  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 SQ SEQUENCE 147 AA; 14994 MW; 34240B0CB3CEB3D6 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 5; Length 147;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 DB 9 ALLVIAL 15

RESULT 48  
 O18679 PRELIMINARY; PRT; 163 AA.  
 ID O18679;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RA Schenkman S., Freitas-Junior L.H.G.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF027879; AAC48356.1; -  
 DR EMBL: AF027878; AAC48355.1; -  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 FT NON\_TER 163 163  
 SQ SEQUENCE 163 AA; 16591 MW; B57ACA32C43CCD33 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 5; Length 163;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 Db 9 ALLVIAL 15

## RESULT 49

ID Q26877 PRELIMINARY; PRT: 165 AA.  
 AC Q26877;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUC.CA-2 PROTEIN.  
 GN MUC.CA-2.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CA1/72;  
 RX MEDLINE=96025798; PubMed=7592617;  
 RA Di Nola J.M., Sanchez D.O., Frasch A.C.;  
 RT "The protozoan Trypanosoma cruzi has a family of genes resembling the  
 RT mucin genes of mammalian cells."  
 RL J. Biol. Chem. 270:24145-24149(1995).  
 DR EMBL; U32346; AAC46904.1;  
 DR InterPro; IPR000458; TRYP\_mucin.  
 DR Pfam; PF01456; TRYP\_mucin; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 165 AA; 17020 MW; 799CBES5EA2148E1 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 165;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 Db 14 ALLVIAL 20

## RESULT 50

ID 09Y596 PRELIMINARY; PRT: 167 AA.  
 AC 09Y596;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE NY-REN-49 ANTIGEN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99438124; PubMed=10508479;  
 RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,  
 RA Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,  
 RA Old L.J.;  
 RT "Antigens recognized by autologous antibody in patients with renal-  
 RT cell carcinoma."  
 RL Int. J. Cancer 83:456-464(1999).  
 DR EMBL; AF151111; AAd42877.1;  
 FT NON\_TER 1  
 SQ SEQUENCE 167 AA; 18956 MW; 43BSA9092232C21 CRC64;

Query Match 6.3%; Score 7; DB 4; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 TISSKD 91  
 Db 119 TISSKD 125

## RESULT 51

ID 09AKW2 PRELIMINARY; PRT: 176 AA.  
 AC 09AKW2;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 20.0 KDA PROTEIN (FRAGMENT).  
 OS Legionella pneumophila.  
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;  
 OC Legionellaceae; Legionella.  
 OX NCBI\_TaxID=446;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SEROGROUP 1, SUBGROUP OJDA;  
 RX MEDLINE=21150471; PubMed=11251842;  
 RA Lueneberg E., Mayer B., Daryab N., Koolstra O., Zaehring U.,  
 RA Rohde M., Swanson J., Frosch M.;  
 RT "Chromosomal insertion and excision of a 30 kb unstable genetic  
 RT element is responsible for phase variation of lipopolysaccharide and  
 RT other virulence determinants in Legionella pneumophila."  
 RL Mol. Microbiol. 39:1259-1271(2001).  
 DR EMBL; AJ277756; CAC33483.1;  
 DR Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 176 AA; 20022 MW; C7E79ADA5AC970BA CRC64;

Query Match 6.3%; Score 7; DB 2; Length 176;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 DUKNTL 109  
 Db 93 DUKNTL 99

## RESULT 52

ID 06I036 PRELIMINARY; PRT: 182 AA.  
 AC 06I036;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 GN EMUCE-31M7.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CL-BRENNER;  
 RX MEDLINE=988225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund I., Sanchez D.O., Frasch A.C.;  
 RA "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions."  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 DR EMBL; AF036426; AAC14235.1;  
 FT NON\_TER 182  
 SQ SEQUENCE 182 AA; 18185 MW; DEAE14AC3BDE2EE CRC64;

Query Match 6.3%; Score 7; DB 5; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 ALLVIAL 12  
 |||||||

Db 9 ALLVAL 15

RESULT 53

ID 061057

PRELIMINARY; PRT; 186 AA.

AC 061057;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE MUCIN-LIKE PROTEIN (FRAGMENT).

GN EMUCP-17.

OS Trypanosoma cruzi.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CL-BRENNER;

RX DI Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;

RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of

RL J. Biol. Chem. 273:10843-10850(1998).

DR EMBL: AF036464; AAC14258.1; -

DR InterPro: IPR000458; TRYP\_mucin.

DR Pfam: PF01456; TRYP\_mucin; 1.

FT NON\_TER 186

SQ SEQUENCE 186 AA; 18663 MW; 986118E03777E82 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 186;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVAL 12

Db 9 ALLVAL 15

RESULT 54

ID 09AH34

PRELIMINARY; PRT; 196 AA.

AC 09AH34;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE AMINOPHENOL OPERON REPRESSOR.

GN MBZR.

OS Pseudomonas putida.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OX NCBI\_TaxID=503;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20100744; PubMed=10633088;

RA Park H.S., Kim H.S.;

RT "Identification and characterization of the nitrobenzene catabolic

RL J. Bacteriol. 182:573-580(2000).

CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL

REGULATORS.

DR EMBL: AF319593; AAK26517.1; -

DR InterPro: IPR000835; HTH\_Matr.

DR Pfam: PF01047; Matr; 1.

DR PRINTS: PRO0347; HTH\_MARR.

DR SMART: SM00347; HTH\_MARR; 1.

KM DNA-binding; Plasmid; Transcription regulation.

SQ SEQUENCE 196 AA; 21885 MW; 6ED6FAEA90A84143 CRC64;

Query Match 6.3%; Score 7; DB 2; Length 196;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 NGLISRV 77

Db 130 NGLISRV 136

RESULT 55

ID 026878

PRELIMINARY; PRT; 197 AA.

AC 026878;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE MUCIN-LIKE PROTEIN.

GN MUC.CA-3.

OS Trypanosoma cruzi.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CA1/72;

RX MEDLINE=96025798; PubMed=7592617;

RA DI Nola J.M., Sanchez D.O., Frasch A.C.;

RT "The protozoan Trypanosoma cruzi has a family of genes resembling the

RL J. Biol. Chem. 270:24146-24149(1995).

DR EMBL: U32447; AAC46906.1; -

DR InterPro: IPR000458; TRYP\_mucin.

DR Pfam: PF01456; TRYP\_mucin; 1.

SQ SEQUENCE 197 AA; 20313 MW; 3270F8704718D157 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 197;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVAL 12

Db 14 ALLVAL 20

RESULT 56

ID 09X302

PRELIMINARY; PRT; 198 AA.

AC 09X302;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE PX01-31.

OS Bacillus anthracis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OX NCBI\_TaxID=1392;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99445483; PubMed=10515943;

RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,

RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,

RA Martinez Y., Rieke D., Svensson R., Jackson P.J.;

RT "Sequence and organization of px01, the large Bacillus anthracis

RL J. Bacteriol. 181:6509-6515(1999).

DR EMBL: AF065404; AAD32335.1; -

KM Plasmid.

SQ SEQUENCE 198 AA; 22859 MW; D76737FD81C43C9D CRC64;

Query Match 6.3%; Score 7; DB 2; Length 198;

Best Local Similarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 40 NELLIDL 46  
 |||||  
 DB 150 NELLIDL 156

## RESULT 57

ID 093RD7 PRELIMINARY; PRT: 198 AA.  
 AC 093RD7;  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ABC MEMBRANE TRANSPORTER HOMOLOGUE (FRAGMENT).  
 OS Brevibacterium fuscum var. dextranilyticum.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Micrococciaceae; Brevibacteriaceae; Brevibacterium.  
 OX NCBI\_Taxid=90748;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mizuno T., Mori H., Nishimoto M., Ito H., Matsui H., Kimura A.,  
 RT "Isolation and sequence of alpha-glucosidase gene from Brevibacterium  
 fuscum var. dextranilyticum strain 0407.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052295; BAB60690.1;  
 FT NON\_TER  
 SQ SEQUENCE 198 AA; 21683 MW; 444C4DC9F9E1F25D0 CRC64;

Query Match 6.3%; Score 7; DB 2; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VLAIVT 15  
 |||||  
 DB 173 VLAIVT 179

## RESULT 58

ID 026885 PRELIMINARY; PRT: 199 AA.  
 AC 026885;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE MUCIN-LIKE GLYCOPROTEIN.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-X10.6;  
 RX MEDLINE-96408676; PubMed-8813683;  
 RA Salazar N.A., Mondragon A., Kelly J.M.;  
 RT "Mucin-like glycoprotein genes are closely linked to members of the  
 trans-sialidase super-family at multiple sites in the Trypanosoma  
 cruzi genome.";  
 RL Mol. Biochem. Parasitol. 78:127-136(1996).  
 DR EMBL; U50160; AAC47164.1;  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 SQ SEQUENCE 199 AA; 19587 MW; 96181D7827E2980 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 DB 14 ALLVIAL 20

RESULT 59  
 ID 026886 PRELIMINARY; PRT: 199 AA.  
 AC 026886;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE MUCIN-LIKE GLYCOPROTEIN.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-X10.6;  
 RX MEDLINE-96408676; PubMed-8813683;  
 RA Salazar N.A., Mondragon A., Kelly J.M.;  
 RT "Mucin-like glycoprotein genes are closely linked to members of the  
 trans-sialidase super-family at multiple sites in the Trypanosoma  
 cruzi genome.";  
 RL Mol. Biochem. Parasitol. 78:127-136(1996).  
 DR EMBL; U50161; AAC47165.1;  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 SQ SEQUENCE 199 AA; 19556 MW; 2429228E4233638E CRC64;

Query Match 6.3%; Score 7; DB 5; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 DB 14 ALLVIAL 20

## RESULT 60

ID 061018 PRELIMINARY; PRT: 200 AA.  
 AC 061018; 061038;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE MUCIN-LIKE PROTEIN.  
 GN EMOCE-1289.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_Taxid=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CL-BRENNER;  
 RX MEDLINE-96225151; PubMed-9556557;  
 RA DI NOLA J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 genes having hypervariable regions.";  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CL-BRENNER;  
 RX D'Orso I., DI NOLA J.M.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF036460; AAC14217.2;  
 DR InterPro: IPR000458; TRYP\_mucin.  
 DR Pfam: PF01456; TRYP\_mucin; 1.  
 SQ SEQUENCE 200 AA; 20352 MW; 8851D29AE5397FB6 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||

Db 10 ALLVIAL 16

RESULT 61

Q9N2S0 PRELIMINARY; PRT; 200 AA.

AC Q9N2S0; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE MUCIN-LIKE PROTEIN.

GN EMUCE-24M16.

OS Trypanosoma cruzi.

OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxId=5693;

RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of genes having hypervariable regions."

RL J. Biol. Chem. 273:10843-10850(1998).

DR EMBL; AF036423; AAF44795.1; -

DR InterPro: IPR000458; TrYP-mucin.

DR Pfam: PF01456; TrYP-mucin; 1.

SO SEQUENCE 200 AA; 20946 MW; D951508C4EA6376E CRC64;

Query Match 6.3%; Score 7; DB 5; Length 200;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12  
Db 10 ALLVIAL 16

RESULT 62

Q9RM08 PRELIMINARY; PRT; 203 AA.

AC Q9RM08; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE HYPOHETICAL 22.0 KDA PROTEIN.

GN DR0608.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI\_TaxId=1299;

RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1."

RL Science 286:1571-1577(1999).

DR EMBL; AE001919; AAF10195.1; -

DR TIGR; DR0608; -

DR InterPro: IPR000182; Acetyltransf\_GCN5.

DR Pfam: PF00583; Acetyltransf; 1.

KW Hypothetical protein; Complete proteome

SO SEQUENCE 203 AA; 21959 MW; 1722A62020C22E7C CRC64;

Query Match 6.3%; Score 7; DB 16; Length 203;

Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GALLVIA 11  
Db 57 GALLVIA 63

RESULT 63

Q9K4K6 PRELIMINARY; PRT; 204 AA.

AC Q9K4K6; 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DE PUTATIVE MEMBRANE TRANSPORT PROTEIN.

GN SCF8.18.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxId=1902;

RT "A set of ordered cosmid and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome."

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL357613; CAB93746.1; -

DR InterPro: IPR001123; LYSE.

DR Pfam: PF01810; LYSE; 1.

SO SEQUENCE 204 AA; 20686 MW; DD1BD686CC04DA1 CRC64;

Query Match 6.3%; Score 7; DB 2; Length 204;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 RYLDGLV 82  
Db 183 RYLDGLV 189

RESULT 64

015772 PRELIMINARY; PRT; 205 AA.

AC 015772; 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DE MUCIN-LIKE PROTEIN (FRAGMENT).

OS Trypanosoma cruzi.

OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI\_TaxId=5693;

RT "Two distinct groups of mucin-like genes are differentially expressed in the developmental stages of Trypanosoma cruzi."

RC STRAIN-Y.

RA FREITAS-Junior L.H., Bioness M.R., Schenkman S.;

SO SEQUENCE 205 AA; 21959 MW; 1722A62020C22E7C CRC64;

RL MOL. Biochem. Parasitol. 93:101-114(1998).  
 DR EMBL: AF027870; AAC48348.1.  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 205  
 SQ SEQUENCE 205 AA; 20673 MW; 9F2383570D523F1C CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 5; Length 205;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

## RESULT 65

ID 015777 PRELIMINARY; PRT; 205 AA.  
 AC 015777;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 ON NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Y;  
 RX MEDLINE=98324409; PubMed=9662032;  
 RA Freitas-Junior L.H., Briones M.R., Schenkman S.;  
 RT "Two distinct groups of mucin-like genes are differentially expressed  
 in the developmental stages of Trypanosoma cruzi";  
 RL Mol. Biochem. Parasitol. 93:101-114(1998).  
 DR EMBL: AF027875; AAC48353.1.  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 205  
 SQ SEQUENCE 205 AA; 20758 MW; 587482D0805F3090 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 5; Length 205;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

## RESULT 66

ID 09Y024 PRELIMINARY; PRT; 205 AA.  
 AC 09Y024;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE MUCIN-LIKE GLYCOPROTEIN PRECURSOR.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 ON NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CL;  
 RA Allen C.L., Mondragon A., Kelly J.M.;  
 RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ239063; CAB43036.1.  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 KW Signal.

FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 205 MUCIN-LIKE GLYCOPROTEIN.  
 SQ SEQUENCE 205 AA; 20697 MW; D5E9424AB1FA185 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 5; Length 205;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

## RESULT 67

ID 061045 PRELIMINARY; PRT; 207 AA.  
 AC 061045;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 ON NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 genes having hypervariable regions";  
 RL J Biol Chem 273:10843-10850(1998).  
 DR EMBL: AF036445; AAC14245.1.  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 SQ SEQUENCE 207 AA; 21108 MW; 8BA4ED7A0052748 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 5; Length 207;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVIAL 12  
 |||||  
 Db 9 ALLVIAL 15

## RESULT 68

ID 061055 PRELIMINARY; PRT; 209 AA.  
 AC 061055;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 ON NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CL-BRENNER;  
 RX MEDLINE=98225151; PubMed=9556557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 genes having hypervariable regions";  
 RL J Biol Chem 273:10843-10850(1998).  
 DR EMBL: AF036462; AAC14256.1.  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 SQ SEQUENCE 209 AA; 21736 MW; DF937DE8518AEB52 CRC64;

```

Query Match
Best Local Similarity 6.3%; Score 7; DB 5; Length 209;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12
    |||||
Db 10 ALLVIAL 16

RESULT 69
Q9Y025 PRELIMINARY; PRT; 210 AA.
ID Q9Y025;
AC 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DE MUCIN-LIKE GLYCOPROTEIN PRECURSOR.
GN MUC3.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CL:
RX MEDLINE-21213828; PubMed-11312580;
RA Allen C.L., Kelly J.M.;
RT "Trypanosoma cruzi: Mucin Pseudogenes Organised in a Tandem Array.";
RL Exp. Parasitol. 97:173-177(2001).
DR EMBL; AJ239064; CAB43037.1; -
DR InterPro: IPR000458; TRYP_mucin.
DR Pfam: PF01456; TRYP_mucin; 1.
KW Signal.
FT SIGNAL. 1 23 POTENTIAL.
FT CHAIN 24 210 MUCIN-LIKE GLYCOPROTEIN.
SQ SEQUENCE 210 AA; 21159 MW; 76F0DBF1EAF83933 CRC64;

Query Match
Best Local Similarity 6.3%; Score 7; DB 5; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12
    |||||
Db 10 ALLVIAL 16

RESULT 70
Q9K2P3 PRELIMINARY; PRT; 215 AA.
ID Q9K2P3;
AC Q9K2P3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE PUTATIVE CDP-DIACYLGLYCEROL-GLYCEROL-3-PHOSPHATE
DE 3-PHOSPHATIDYL-TRANSFERASE.
GN SCL18A.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetidae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

```

---

```

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL353861; CAB88885.1; -
DR InterPro: IPR000462; CDP-OH_P_transf.
DR PROSITE: PS00379; CDP_ALCOHOL_P_TRANSF. 1.
SQ SEQUENCE 215 AA; 23771 MW; 78923574A134FCF6 CRC64;

Query Match
Best Local Similarity 6.3%; Score 7; DB 2; Length 215;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12
    |||||
Db 62 ALLVIAL 68

RESULT 71
Q26880 PRELIMINARY; PRT; 215 AA.
ID Q26880;
AC Q26880;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE MUCIN-LIKE PROTEIN.
GN MUC. RA-2.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RA:
RX MEDLINE-96025798; PubMed-7592617;
RA Di Nola J.M., Sanchez D.O., Frasch A.C.;
RT "The protozoan Trypanosoma cruzi has a family of genes resembling the
RT mucin genes of mammalian cells.";
RL J. Biol. Chem. 270:24146-24149(1995).
DR EMBL; U32449; AAC46908.1; -
DR InterPro: IPR000458; TRYP_mucin.
DR Pfam: PF01456; TRYP_mucin; 1.
SQ SEQUENCE 215 AA; 21819 MW; 980A8B1D1AE56F3A CRC64;

Query Match
Best Local Similarity 6.3%; Score 7; DB 5; Length 215;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12
    |||||
Db 9 ALLVIAL 15

RESULT 72
Q962W6 PRELIMINARY; PRT; 216 AA.
ID Q962W6;
AC Q962W6;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DE MUCIN-LIKE PROTEIN MOC-LOC2.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CL-BRENNER;
RA Di Nola J.M., Frasch A.C.C.;

```

RT "Hypervariable mucin genes from Trypanosoma cruzi."  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF398551; AAK94014.1; -  
SO SEQUENCE 216 AA; 21815 MW; 01C85738541B86C6 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 5; Length 216;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 ALLVIAL 12  
DB 10 ALLVIAL 16

RESULT 73  
Q9FRK3 PRELIMINARY; PRT; 216 AA.  
ID Q9FRK3  
AC Q9FRK3  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 23.8 KDA PROTEIN.  
GN OSJNB0064P21.16.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAINCV; NIPONBARE;  
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
RA Zisman V., Pal G., Bowman G.L., Fujii C.Y., Vanaken S.E.,  
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
RT "Oryza sativa chromosome 10 [BAC OSJNB0064P21 genomic sequence.";  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC073166; AAG46103.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 216 AA; 23796 MW; F9D49F6A9CC224FC CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 10; Length 216;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 LLYVALL 13  
DB 10 LLYVALL 16

RESULT 74  
Q9CMX6 PRELIMINARY; PRT; 230 AA.  
ID Q9CMX6  
AC Q9CMX6  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN PM0674.  
GN PM0674.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;  
OC Pasteurella.  
OX NCBI\_TaxID=147;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PM70;  
RA MEDLINE-21145866; PubMed-11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida PM70.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
DR EMBL; AE006103; AAK02758.1; -

KW Hypothetical protein: Complete proteome  
SQ SEQUENCE 230 AA; 25287 MW; 258CA22164678F62 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 16; Length 230;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 14 VTQALGV 20  
DB 39 VTQALGV 45

RESULT 75  
Q9RWM1 PRELIMINARY; PRT; 244 AA.  
ID Q9RWM1  
AC Q9RWM1  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 26.2 KDA PROTEIN.  
GN DR0554.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RI;  
RA MEDLINE-20036896; PubMed-10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans RI.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001913; AAP10136.1; -  
DR TIGR; DR0554; -  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 244 AA; 26169 MW; 5045FE7B01986AAE CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 16; Length 244;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 LLYVALL 13  
DB 54 LLYVALL 60

RESULT 76  
Q9NTB9 PRELIMINARY; PRT; 253 AA.  
ID Q9NTB9  
AC Q9NTB9  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL 27.5 KDA PROTEIN.  
GN DRP2P43J212.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-TESTIS;  
RA Ansong W., Winkler U., Meyers H.W., Weil B., Wiemann S.;  
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL137395; CAB70723.1; -



DR InterPro: IPR000372; LRR\_Nterm.  
 DR SMART: SM00013; LRRNT; 1.  
 KW Hypothetical protein  
 SQ SEQUENCE 253 AA; 27499 MW; 286D4E87F6625D56 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 4; Length 253;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVVLALL 13  
 |||||  
 Db 167 LVVLALL 173

RESULT 77  
 O9KW46 PRELIMINARY; PRT; 261 AA.  
 AC O9KW46;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE COMPONENT OF TYPE IV SECRETION SYSTEM.  
 GN VIRB9.  
 OS Wolbachia sp. wTat.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Wolbachieae; Wolbachia.  
 OX NCBI\_TaxID=110061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WTAT;  
 RX MEDLINE=20507828; PubMed=11053403;  
 RA Masui S., Sasaki T., Ishikawa H.;  
 RT "Genes for the Type IV Secretion System in an Intracellular Symbiont,  
 RT Wolbachia, a Causative Agent of Various Sexual Alterations in  
 RT Arthropods.";  
 RL J. Bacteriol. 182:6529-6531(2000).  
 DR EMBL: AB045234; BAA97433.1; -  
 SQ SEQUENCE 261 AA; 30373 MW; 563328F948552F38 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 2; Length 261;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVVLALL 13  
 |||||  
 Db 5 LVVLALL 11

RESULT 78  
 O92C38 PRELIMINARY; PRT; 262 AA.  
 AC O92C38;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDSA PROTEIN  
 GN CDSA OR LIN1353.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacilli/Staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchand E., Durant L., Dussurget O.,  
 RA Entlan K.-D., Fslhi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,

RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkai G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schueller T., Simoes N., Tietz A.,  
 RA Vazquez-Boiland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL: AL596168; CAC96584.1; -  
 DR LISTLIST; LIN01353; -  
 KW Complete proteome.  
 SQ SEQUENCE 262 AA; 29097 MW; 9CA40332625152AA CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 16; Length 262;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LVVLALL 14  
 |||||  
 Db 201 LVVLALL 207

RESULT 79  
 O9X8G2 PRELIMINARY; PRT; 266 AA.  
 AC O9X8G2;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.  
 GN SCF7.05C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RT Submitted (May-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rastandream M.A.;  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=843436;  
 RA Redenbach M., Kieser H.M., Denapalae D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL049819; CAB42664.1; -  
 DR InterPro: IPR002106; AA\_TRNA\_Ligase\_II.  
 DR PROSITE: PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 SQ SEQUENCE 266 AA; 28200 MW; A65858A932E2D590 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 2; Length 266;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVLAL 12  
 |||||  
 Db 250 ALLVLAL 256

RESULT 80  
 O22798 PRELIMINARY; PRT; 268 AA.  
 AC O22798;  
 ID O22798

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE CUC3 PROTEIN (PUTATIVE NAM (NO APICAL MENIStEM)-LIKE PROTEIN)  
 DE (AF033480/FA09.25)  
 GN FAP.25 OR AF033480.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, COLDWIA;  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Theologis A., Ecker J. Davis R.W.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bower L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin X., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002332; AAB80665.1;  
 DR EMBL: AF325080; AAK17148.1;  
 DR EMBL: AF410299; AAK95285.1;  
 DR InterPro: IPR003441; NAM.  
 DR Pfam: PF02365; NAM; 1;  
 SO SEQUENCE 268 AA; 31124 MM; 02F34981BA78C336 CRC64;

Query Match 6.38; Score 7; DB 10; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 25 TCPIFYD 31  
 DB 197 TCPIFYD 203

RESULT 81  
 O962H7 PRELIMINARY; PRT; 276 AA.  
 AC O962H7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE MUCIN EMUCT-4.  
 GN EMUCT-4.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=9825151; PubMed=9555557;  
 RA Di Nola J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;  
 RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of  
 RT genes having hypervariable regions."  
 RL J. Biol. Chem. 273:10843-10850(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Di Nola J.M., Frasch A.C.;  
 RT "mucin mucin-like gene group II."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF032682; AAK49789.1;  
 SO SEQUENCE 276 AA; 27336 MM; 2A8A7BB1F9EA3FBD CRC64;

Query Match 6.38; Score 7; DB 5; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 ALLVIAL 12  
 DB 9 ALLVIAL 15

RESULT 82  
 O15771 PRELIMINARY; PRT; 277 AA.  
 AC O15771;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MUCIN-LIKE PROTEIN (FRAGMENT).  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y;  
 RX MEDLINE=98324409; PubMed=9662032;  
 RA Freitas-Junior L.H., Briones M.R., Schenkman S.;  
 RT "Two distinct groups of mucin-like genes are differentially expressed  
 RT in the developmental stages of Trypanosoma cruzi."  
 RL MOL. Biochem. Parasitol. 93:101-114(1998).  
 DR EMBL: AF027869; AAC48347.1;  
 DR InterPro: IPR000458; Tryp\_mucin.  
 DR Pfam: PF01456; Tryp\_mucin; 1.  
 FT NON\_TER 277  
 SO SEQUENCE 277 AA; 27403 MM; F1FB3BB0B121751 CRC64;

Query Match 6.38; Score 7; DB 5; Length 277;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 6 ALLVIAL 12  
 DB 9 ALLVIAL 15

RESULT 83  
 P73885 PRELIMINARY; PRT; 279 AA.  
 AC P73885;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 31.2 KDA PROTEIN.  
 GN STR0267.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97061201; PubMed=8905231;  
 RA Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions."  
 RL DNA Res. 3:109-116(1996).  
 DR EMBL: D90910; BAA17949.1; -  
 DR Interpro: IPR002917; MMR\_HSR1.  
 DR Pfam: PF01926; MMR\_HSR1; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 279 AA; 31201 MW; 9AA60F311BD77763 CRC64;

Query Match 6.3%; Score 7; DB 16; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LLLDLSL 48  
 |||||  
 Db 213 LLLDLSL 219

RESULT 84  
 064761  
 ID 064761 PRELIMINARY; PRT; 303 AA.  
 AC 064761;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE PUTATIVE PHOSPHATIDYLINOSITOL-GLYCAN SYNTHASE.  
 GN APTG34980.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Fossidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxId=3702;

RT [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA MEDLINE-20083487; PubMed-10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlynn T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhagen G.P., Preuss D., Niemman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT \*Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RL thaliana 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004238; AXC12837.1; -  
 SQ SEQUENCE 303 AA; 34200 MW; 88E6BF82F4940C92 CRC64;

Query Match 6.3%; Score 7; DB 10; Length 303;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 LLLDLSL 48  
 |||||  
 Db 89 LLLDLSL 95

RESULT 85  
 09AVAN9  
 ID 09AVAN9 PRELIMINARY; PRT; 322 AA.  
 AC 09AVAN9;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CG14511 PROTEIN.  
 GN CG14511.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;

RT [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos B.L.G.,  
 RA Abril J.F., Adganyi A., An H.-J., Andrews-Prankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003768; AAF56864.1; -  
 DR FlyBase: FBgn0039641; CG14511.  
 SQ SEQUENCE 322 AA; 36056 MW; 94F48488E19A1F29 CRC64;

Query Match 6.3%; Score 7; DB 5; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVALL 12  
 |||||  
 Db 169 ALLVALL 175

RESULT 86  
 09AYV5  
 ID 09AYV5 PRELIMINARY; PRT; 322 AA.  
 AC 09AYV5;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MINOR STRUCTURAL PROTEIN 4.  
 OS Lactococcus lactis bacteriophage Tuc2009.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxId=35241;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA van Sinderen D., van de Gucht M., Seegers J.F.M.L., Fitzgerald G.F.,  
 RT "Molecular analysis of the temperate lactococcal phage Tuc2009."  
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF109874; AAK19876.1; -  
 SO SEQUENCE 322 AA; 36399 MW; 9FCAD2A50819292 CRC64;

Query Match 6.3%; Score 7; DB 9; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 38 NGNELL 44  
 Db 268 NGNELL 274

RESULT 87  
 ID 091523 PRELIMINARY; PRT: 322 AA.  
 AC 091523;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE HYPOTHETICAL PROTEIN PA0543  
 GN PA0543.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID-287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE-2043737; PubMed-10984043;  
 RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004490; AAG03932.1;  
 DR InterPro: IPR000379; Est\_11b-thioest-actsite.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 322 AA; 35138 MW; 0DD1307E241E74AA CRC64;

Query Match 6.3%; Score 7; DB 16; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 LVALLV 13  
 Db 14 LVALLV 20

RESULT 88  
 ID 09N282 PRELIMINARY; PRT: 324 AA.  
 AC 09N282;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE MMP-9 (FRAGMENT).  
 GN MMP-9.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID-9913;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Sato T., Hirata M., Ito A., Hashizume K.;  
 RT "Bovine endometrial MMP-9 mRNA."  
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB043996; BAA96389.1; -  
 SO SEQUENCE 324 AA; 36273 MW; 4137C9820C28E080 CRC64;

Query Match 6.3%; Score 7; DB 6; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 LVALLV 14  
 Db 7 LVALLV 13

RESULT 89  
 ID 09WY83 PRELIMINARY; PRT: 341 AA.  
 AC 09WY83;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 39.6 KDA PROTEIN.  
 GN TM0241.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID-2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MSB8 / DSM 3109;  
 RX MEDLINE-99287316; PubMed-10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,  
 RA Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of Thermotoga maritima."  
 RL Nature 399:323-329(1999).  
 DR EMBL: AE001707; AAD35332.1; -  
 DR TIGR: TM0241;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 341 AA; 39588 MW; 60C36B8A551496CD CRC64;

Query Match 6.3%; Score 7; DB 16; Length 341;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 LVALLV 14  
 Db 322 LVALLV 328

```

RESULT 90
P73865 PRELIMINARY: PRT; 350 AA.
AC P73865;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SENSOR TRANSDUCTION HISTIDINE KINASE.
GN SLI1590.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
CC DNA Res. 3:109-136(1996).
CC -I SIMILARITY: TO PROKARYOTE SENSOR TRANSDUCTION PROTEINS.
DR EMBL: D90910; BAI17927.1; -.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR003661; His_KinA.
DR InterPro: IPR004359; His_Kin_s1g.
DR Pfam: PF02518; HATPase_C; 1.
DR Pfam: PF00512; signal; 1.
DR SMART: SM00387; HATPase_C; 1.
DR SMART: SM00388; HSKA; 1.
KW Complete proteome; kinase; phosphorylation; sensory transduction;
KW Transferrase.
SQ SEQUENCE 350 AA; 39445 MW; 5730D3D7DA08F5B CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 VLAIVT 15
DB 86 VLAIVT 92

RESULT 91
042853 PRELIMINARY: PRT; 355 AA.
AC 042853;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PEROXIDASE BP 2A PRECURSOR (EC 1.11.1.7).
GN PRX5.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92380512; PubMed=1355062;
RA Mellade B., Sasussen S.K.;
RT "Structure and chromosomal localization of the gene encoding barley
RT seed peroxidase BP 2A.";
CC Gene 118:261-266(1992).
DR EMBL: M83611; AAA32974.1; -.
DR HSSA: P00443; 2ATJ.
DR InterPro: IPR002016; Peroxidase.
DR Pfam: PF00141; peroxidase; 1.

```

```

DR PRINTS: PR00458; PEROXIDASE.
DR PROSITE: PS00435; PEROXIDASE.1; 1.
KW Signal; Oxidoreductase; Peroxidase.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 355 PEROXIDASE BP 2A.
SQ SEQUENCE 355 AA; 38117 MW; CB520193E73538B2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 355;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALLVIAL 12
DB 13 ALLVIAL 19

RESULT 92
0970V4 PRELIMINARY: PRT; 360 AA.
AC 0970V4;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE AMINOTRANSFERASE, CLASS-V.
GN SP1094.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Debay R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
CC Science 293:498-506(2001).
DR EMBL: AE007410; AAK75206.1; -.
DR TIGR: SP1094; -.
DR InterPro: IPR000192; Aminotransf_class_V.
DR Pfam: PF00266; aminotran_5; 1.
DR PROSITE: PS00595; AA_TRANSFR_CLASS_5; 1.
KW Transferrase; Complete proteome.
SQ SEQUENCE 360 AA; 39572 MW; E88EBBA9D9A8128F CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 360;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 QNTVEDL 104
DB 341 QNTVEDL 347

RESULT 93
09BXp9 PRELIMINARY: PRT; 362 AA.
AC 09BXp9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE TANDEM FYVE FINGERS-1 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN  
 [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=21157035; PubMed=11256955;  
 RA C. F. Cheung P., Trinkle-Mulcahy L., Cohen P., M. Lucocq J.;  
 "Characterization of a novel phosphatidylinositol 3-phosphate-binding  
 protein containing two FYVE fingers in tandem that is targeted to the  
 Golgi.";  
 RT Biochem. J. 355:113-121(2001);  
 DR EMBL: AF311602; AAK27339.1; -;  
 DR InterPro: IPR000306; Znf\_FYVE.  
 DR Pfam: PF01363; FYVE; 2.  
 DR SMART: SM00064; FYVE; 2.  
 SO SEQUENCE 362 AA; 40260 MW; 70B405CA16F3BBD CRC64;

Query Match 6.3%; Score 7; DB 4; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 GEAVONT 100  
 DB 267 GEAVONT 273

RESULT 94  
 ID 042854 PRELIMINARY; PRT; 364 AA.  
 AC 042854;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE BP 28 (EC 1.11.1.7).  
 GN Pex6.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NK1558;  
 RA Theilade B., Rasmussen S.K.;  
 RT Submitted (JMB-1993) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: Z23131; CA80667.1; -;  
 DR HSSP: P22195; 1SCH  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PR00141; peroxidase\_1;  
 DR PRINTS: PR00458; peroxidase\_1;  
 DR PROSITE: PS00435; peroxidase\_1; 1.  
 KW Oxidoreductase; Peroxidase;  
 SO SEQUENCE 364 AA; 39114 MW; BD1BF36EB2ADFE2C CRC64;

Query Match 6.3%; Score 7; DB 10; Length 364;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 ALLVLTAL 12  
 DB 13 ALLVLTAL 19

RESULT 95  
 ID 09KMW6 PRELIMINARY; PRT; 374 AA.  
 AC 09KMW6;  
 DT 01-OCT-2000 (TEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN VCA0200;

GN VCA0200  
 OS Vibrio cholerae  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 NCBI\_TaxID=666;  
 RN  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SPROTYPE 01;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gailin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey F.K., Peterson J.D., Richardson D.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson S.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi C., White P.,  
 RA McDonald L., Ueberlack T., Fleischmann R.D., Nierns W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RT Nature 406:477-483(2000).  
 DL EMBL: AE004360; AAF96113.1; -;  
 DR TIGR: VCA0200; -;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 374 AA; 42042 MW; 337EDA396DDP5F7 CRC64;

Query Match 6.3%; Score 7; DB 16; Length 374;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 VLVLTAL 15  
 DB 254 VLVLTAL 260

RESULT 96  
 ID 0914M7 PRELIMINARY; PRT; 385 AA.  
 AC 0914M7;  
 DT 01-MAR-2001 (TEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last annotation update)  
 DE PROBABLE MFS TRANSPORTER.  
 GN PA1108.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O., T. Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brady L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RT Nature 406:959-964(2000).  
 DL EMBL: AE004541; AAG04497.1; -;  
 DR InterPro: IPR001958; TCR\_Teta.  
 DR PRINTS: PR01035; TCR\_Teta.  
 KW Complete proteome.  
 SO SEQUENCE 385 AA; 40554 MW; BDE11D1D15206944 CRC64;

Query Match 6.3%; Score 7; DB 16; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 LVLVLTAL 13  
 DB 10 LVLVLTAL 16

```

RESULT 97
ID 007417 PRELIMINARY; PRT; 390 AA.
AC 007417;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN HYPOPHOSPHATE 42.2 KDa PROTEIN.
LN LPRK OR RV0173 OR MTC128.13.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badock K., Besham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
DR EMBL; 297050; CAB09756.1; -.
DR Tuberculist; RV0173; -.
DR InterPro; IPR003399; Mce.
DR Pfam; PF02470; mce; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 390 AA; 42239 MW; 0B6929A319D15D23 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 390;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 7 LVLLALL 13
ID 09X7S6 PRELIMINARY; PRT; 400 AA.
AC 09X7S6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN PUTATIVE MEMBRANE PROTEIN.
LN SC5H1.18.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);

```

```

RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kiese H.M., Denaplatte D., Eichner A., Cullum J.,
RA Kinsashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96 (1996).
DR EMBL; AL049863; CAB42943.1; -.
SQ SEQUENCE 400 AA; 41458 MW; C0572F19D31F23EE CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 400;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 8 LVLLALL 14
ID 09NRG5 PRELIMINARY; PRT; 403 AA.
AC 09NRG5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
GN HNYA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER CANCER TISSUE;
RT Li Y., Wu T., Xu S., Ren S., Chen Z., Han Z.;
RT "A novel gene expressed in human liver cancer tissue."
DR EMBL; AF226052; ANF6952.1; -.
SQ SEQUENCE 403 AA; 47347 MW; 28CBD97B4114819F CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 403;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 85 TISSKD 91
ID 0914K4 PRELIMINARY; PRT; 422 AA.
AC 0914K4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN PROBABLE MFS TRANSPORTER.
LN PAL131.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

```

RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL: AE004543; AAC04520.1; -;  
DR InterPro: IPR003662; sub\_transporter.  
DR PRINTS: PR01035; TCRTFA.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Complete proteome; Transmembrane.  
SQ SEQUENCE 422 AA; 44365 MW; AB246ABE616A50D2 CRC64;

Query Match 6.38; Score 7; DB 16; Length 422;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GALLVLA 11  
DB 397 GALLVLA 403

Search completed: July 9, 2002, 16:19:46  
Job time: 96 sec